

HUMBOLDT-UNIVERSITÄT ZU BERLIN  
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN  
ZUR BIBLIOTHEKS- UND  
INFORMATIONSWISSENSCHAFT

HEFT 461

OBJEKTBSCHREIBENDE METADATEN IN DER DEUTSCHEN  
DIGITALEN BIBLIOTHEK

ANALYSE DER EIGNUNG ALS INFORMATIONSFILTER IM  
RETRIEVAL

VON  
FRANCESCA SCHULZE



OBJEKTBESCHREIBENDE METADATEN IN DER DEUTSCHEN  
DIGITALEN BIBLIOTHEK

ANALYSE DER EIGNUNG ALS INFORMATIONSFILTER IM  
RETRIEVAL

VON  
FRANCESCA SCHULZE

---

Berliner Handreichungen zur  
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn  
Herausgegeben von  
Vivien Petras  
Humboldt-Universität zu Berlin

Heft 461

**Schulze, Francesca**

Objektbeschreibende Metadaten in der Deutschen Digitalen Bibliothek : Analyse der Eignung als Informationsfilter im Retrieval / von Francesca Schulze. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2020. - 117 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 461)

ISSN 14 38-76 62

#### Abstract:

Diese Arbeit geht von der Beobachtung aus, dass die Retrievalmöglichkeiten in der Deutschen Digitalen Bibliothek (DDB) nicht befriedigend sind. Es fehlen vor allem navigierbare Sucheinstiege und wirksame Filtermöglichkeiten. Deshalb werden im ersten Teil dieser Arbeit Vorschläge für drei Informationsfilter entwickelt, die spartenübergreifend fundamentale Facetten abdecken und ein nutzerfreundliches Browsen im DDB-Portal ermöglichen könnten: »Sachkategorie«, »Themenschlagwort« und »Objektgattung«. Die erfolgreiche Anwendung dieser Filter setzt jedoch voraus, dass bestimmte Metadatenelemente in den gelieferten Objekten vorhanden und mit geeigneten Werten belegt sind. Im empirischen Teil dieser Arbeit wird deshalb untersucht, ob und wenn ja inwieweit die Metadaten der gelieferten Objekte in der DDB die Voraussetzungen für ein filterbasiertes Retrieval mitbringen.

Die Untersuchung geht von zwei Fragen an die Metadaten der DDB-Objekte aus: 1. Definieren die einzelnen Metadatenschemata übereinstimmend Elemente, die sich spartenübergreifend als Filter eignen? 2. Eignen sich die Inhalte der Metadatenelemente als Navigationselemente?

Das Erstellen von Konkordanzen zwischen den Schemata zeigte, dass die entsprechenden Elemente semantisch weitgehend korrespondieren und nahtlos auf die vorgeschlagenen drei Filter abgebildet werden konnten. Die Untersuchung einer randomisierten Stichprobe aus 20.000 Datensätzen über die Grundgesamtheit von etwa 20 Millionen Datensätzen in der DDB ergibt, dass die objektbeschreibenden Metadaten die genannten Voraussetzungen für ein filterbasiertes Retrieval in einigen Aspekten nicht erfüllen. Im abschließenden Teil der Arbeit werden daher Überlegungen angestellt, welche Maßnahmen zur Verbesserung der Metadaten eingeleitet werden müssten, damit die vorgeschlagenen Informationsfilter in der DDB sinnvoll eingesetzt werden können.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im weiterbildenden Masterstudiengang im Fernstudium Bibliotheks- und Informationswissenschaft (Library and Information Science, M. A. (LIS)) an der Humboldt- Universität zu Berlin.

Eine Online-Version ist auf dem edoc Publikationsserver der Humboldt-Universität zu Berlin verfügbar.



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) Lizenz.

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis .....</b>	<b>5</b>
<b>Abbildungsverzeichnis .....</b>	<b>7</b>
<b>Tabellenverzeichnis .....</b>	<b>8</b>
<b>Abkürzungsverzeichnis .....</b>	<b>9</b>
<b>1     Einleitung.....</b>	<b>10</b>
1.1    Problemstellung und Motivation.....	10
1.2    Forschungsfrage und Untersuchungsziel .....	12
1.3    Vorgehensweise und thematische Eingrenzung .....	13
1.4    Bezeichnungen und Definitionen .....	14
1.5    Typografische Gestaltung, orthografische und sprachliche Konventionen .....	15
<b>2     Voruntersuchung .....</b>	<b>16</b>
2.1    Suche in der Deutschen Digitalen Bibliothek .....	16
2.2    Informationsfilter in Webanwendungen .....	25
2.3    Mögliche Informationsfilter für die DDB .....	34
2.4    Mögliche kontrollierte Vokabulare für die vorgeschlagenen Filter .....	36
<b>3     Untersuchungsmethode .....</b>	<b>48</b>
3.1    Erhebung einer Zufallsstichprobe aus den Lieferdaten .....	48
3.2    Auswahl der Elemente aus den Metadatenschemata .....	52
3.3    Untersuchungskriterien .....	55
3.4    Aufbereitung der Forschungsdaten.....	56
<b>4     Stichprobenanalyse nach Metadatenformat .....</b>	<b>58</b>
4.1    MAchine-Readable Cataloging (MARC) .....	58
4.2    Metadata Object Description Schema (MODS) .....	62
4.3    Encoded Archival Description (EAD).....	67
4.4    Lightweight Information Describing Objects (LIDO).....	70
4.5    Dublin Core Metadata Element Set (DCMES).....	76
<b>5     Fazit und Ausblick .....</b>	<b>79</b>
5.1    Formatübergreifende Zusammenfassung der Ergebnisse .....	79
5.2    Überlegungen für das weitere Vorgehen zur Verbesserung der Retrievalmöglichkeiten in der DDB.....	81

<b>Literatur und Quellen .....</b>	<b>88</b>
Literaturverzeichnis.....	88
Dokumente der Deutschen Digitalen Bibliothek .....	92
Metadatenstandards, Schemata, Anwendungsprofile .....	92
Vokabularstandards, Regelwerke und Anwendungsleitfäden zur Indexierung .....	94
Einzelne Vokabulare, Schlagwortlisten, Codelisten .....	95
<b>Forschungsdatenverzeichnis .....</b>	<b>98</b>
<b>Anhang A: Zusammenstellung der untersuchten Metadatenelemente .....</b>	<b>100</b>
MAchine-Readable Cataloging, 1999 Edition .....	100
Metadata Object Description Schema, MODS 3.6.....	104
Encoded Archival Description, Version EAD3 .....	108
Lightweight Information Describing Objects 1.0 .....	110
Dublin Core Metadata Element Set, Version 1.1 .....	113
<b>Anhang B: Übersicht der in der DDB verwendeten kontrollierten Vokabulare .....</b>	<b>115</b>

## Abbildungsverzeichnis

Abbildung 2.1: Startseite des DDB-Portals .....	18
Abbildung 2.2: Ergebnisse für Suchanfrage nach ›malerei‹ in der DDB .....	19
Abbildung 2.3: Werte mit Bestandteil ›ök‹ im Stichwortfilter der DDB .....	20
Abbildung 2.4: Suchvorschläge für ›malerei‹ im Freitextsuchfeld der DDB .....	21
Abbildung 2.5: Ergebnisse für Suchanfrage nach ›ölmalerei‹ in der DDB .....	22
Abbildung 2.6: Detailseite für das Objekt »Die Malerei« in der DDB .....	23
Abbildung 2.7: Informationsverlust bei der Suche im DDB-Portal am Beispiel ›Malerei‹ .....	25
Abbildung 2.8: »Stöbern in Kategorien« bei ebay .....	28
Abbildung 2.9: Suchvorschläge mit Kategorien bei ebay .....	29
Abbildung 2.10: Aktivierbare Filter bei ebay .....	29
Abbildung 2.11: Sucheinstieg nach »Sammlungen« in Europeana Collections.....	30
Abbildung 2.12: Einstiegseite der Sammlung »Europeana Art«.....	31
Abbildung 2.13: Detailseite für das Objekt »Die Malerei«in Europeana Collections .....	32
Abbildung 2.14: Ergebnisseite für das Navigationselement ›Malerei‹ in Europeana Collections	33
Abbildung 2.15: Begriffsleiter für die Klasse ›751.45‹ in der »DDC Deutsch« .....	41
Abbildung 2.16: Sachbegriff ›Ölmalerei‹ im Katalog der Deutschen Nationalbibliothek.....	44
Abbildung 2.17: Begriffsleiter für den AAT-Deskriptor ›oil paintings (visual works)‹ .....	46
Abbildung 2.18: Suchausweitung über GND-Hierarchierelationen am Beispiel ›Malerei‹ .....	47
Abbildung 2.19: Suchausweitung über DDC- und GND-Sachgruppen am Beispiel ›Malerei‹ .....	47
Abbildung 3.1: Verteilung der Datenformate in der Stichprobe (22.02.2017).....	51
Abbildung 3.2: Verteilung der Datenformate in der Grundgesamtheit (15.05.2018).....	51

## Tabellenverzeichnis

Tabelle 3.1: Übersicht der Datensätze in der Stichprobe nach Metadatenformat .....	50
Tabelle 3.2: Metadatenelemente für den vorgeschlagenen Filter »Sachkategorie« .....	52
Tabelle 3.3: Metadatenelemente für den vorgeschlagenen Filter »Themenschlagwort« .....	53
Tabelle 3.4: Metadatenelemente für den vorgeschlagenen Filter »Objektgattung« .....	54
Tabelle 3.5: Elementinstanzen in der Tabelle »mods_stichprobe« .....	57
Tabelle 3.6: Elementinstanzen in der Tabelle »mods_classification« .....	57
Tabelle 4.1: Datenauswertung der MARC-Elemente für Notationen .....	59
Tabelle 4.2: Kontrollierte Vokabulare in den MARC-Elementen für Notationen .....	60
Tabelle 4.3: Datenauswertung des MARC-Elementes »Sachschlagwort« .....	60
Tabelle 4.4: Kontrolliertes Vokabular im MARC-Element »Sachschlagwort« .....	61
Tabelle 4.5: Datenauswertung des MARC-Elementes »Genre/Formschlagwort« .....	61
Tabelle 4.6: Kontrollierte Vokabulare im MARC-Element »Genre-/Formschlagwort« .....	62
Tabelle 4.7: Datenauswertung des MODS-Elementes »Klassifikation« .....	63
Tabelle 4.8: Kontrollierte Vokabulare im MODS-Element »Klassifikation« .....	64
Tabelle 4.9: Datenauswertung des MODS-Elementes »Sachschlagwort« .....	64
Tabelle 4.10: Kontrollierte Vokabulare im MODS-Element »Sachschlagwort« .....	65
Tabelle 4.11: Datenauswertung für das MODS-Element »Gattungsbegriff« .....	65
Tabelle 4.12: Kontrollierte Vokabulare im MODS-Element »Gattungsbegriff« .....	67
Tabelle 4.13: Datenauswertung für das EAD-Element »Gegenstand« .....	68
Tabelle 4.14: Datenauswertung für das EAD-Element »Genre- und Formbegriffe« .....	69
Tabelle 4.15: Kontrolliertes Vokabular im EAD-Element »Genre- und Formbegriffe« .....	70
Tabelle 4.16: Datenauswertung für das LIDO-Element »Classification« .....	71
Tabelle 4.17: Kontrollierte Vokabulare im LIDO-Element »Classification« .....	72
Tabelle 4.18: Datenauswertung für das LIDO-Element »Subject Concept« .....	73
Tabelle 4.19: Kontrollierte Vokabulare im LIDO-Element »Subject Concept« .....	74
Tabelle 4.20: Datenauswertung für das LIDO-Element »Object/Work Type« .....	75
Tabelle 4.21: Kontrollierte Vokabulare im LIDO-Element »Object/Work Type« .....	76
Tabelle 4.22: Datenauswertung für das DCMES-Element »Thema« .....	77
Tabelle 4.23: Kontrolliertes Vokabular für das DCMES-Element »Thema« .....	77
Tabelle 4.24: Datenauswertung für das DCMES-Element »Typ« .....	78
Tabelle 4.25: Kontrolliertes Vokabular im DCMES-Element »Typ« .....	78



## Abkürzungsverzeichnis

<b>AAT</b>	Art & Architecture Thesaurus
<b>BAM</b>	Bibliotheken, Archive, Museen
<b>DCMES</b>	Dublin Core Metadata Element Set
<b>DDB</b>	Deutsche Digitale Bibliothek
<b>DDC</b>	Dewey-Dezimalklassifikation
<b>DFG</b>	Deutsche Forschungsgemeinschaft
<b>DNB</b>	Deutsche Nationalbibliothek
<b>EAD</b>	Encoded Archival Description
<b>ESE</b>	Europeana Semantic Elements
<b>GBV</b>	Gemeinsamer Bibliotheksverbund
<b>GND</b>	Gemeinsame Normdatei
<b>GND O</b>	GND-Ontologie
<b>LCSH</b>	Library of Congress Subject Headings
<b>LIDO</b>	Lightweight Information Describing Objects
<b>LOD</b>	Linked Open Data
<b>MARC</b>	MAchine-Readable Cataloging
<b>MODS</b>	Metadata Object Description Schema
<b>OBG</b>	Oberbegriffsdatei
<b>OCLC</b>	Online Computer Library Center
<b>RSWK</b>	Regeln für die Schlagwortkatalogisierung
<b>SKOS</b>	Simple Knowledge Organization System
<b>SWD</b>	Schlagwortnormdatei
<b>RDF</b>	Resource Description Framework
<b>URI</b>	Uniform Resource Identifier
<b>URL</b>	Uniform Resource Locator
<b>WWW</b>	World Wide Web

# 1 Einleitung

## 1.1 Problemstellung und Motivation

George Oates, ehemalige Chefredakteurin bei Flickr, sprach am Rande der EuropeanaTech-Konferenz 2015 in Paris aus, was seit dem Beta-Launch der Europeana im Jahr 2009 als generelles Problem der aus öffentlicher Hand geförderten digitalen Kulturerbeportale gesehen wird:

Wenn Sie an eine tolle Sammlung von Kulturerbe denken und Ihre erste Anlaufstelle ein Suchfeld ist, ist das einfach nur tragisch. [...] Dinge wie Google sind gut, wenn Sie überall nach allem suchen wollen, aber in unserer Welt wissen wir, um welche Art von Inhalt es geht, also sollten wir diesem Inhalt [...] ein wenig Aufmerksamkeit schenken. (Pekel, 2015b)

Kulturerbeportale wie Europeana oder die Deutsche Digitale Bibliothek (DDB) bieten Wissenschaftlern und Studierenden, Lehrenden und Lernenden und allen Kulturinteressierten den Vorteil, verteilte Sammlungen aus den verschiedenen Sparten kultureller Einrichtungen wie Bibliothek, Archiv und Museum unter einem Dach durchforschen zu können. Dabei spielen gute Suchmöglichkeiten eine wichtige Rolle.

Im Unterschied zu Suchen im World Wide Web (WWW) sind die Erwartungen an Suchen in Kulturerbeportalen jedoch andere: Während ein Google-Nutzer bei einer informationsorientierten Suche<sup>1</sup> ohne Erwartung an eine themenbezogene Vollständigkeit eine Anfrage in das Suchfeld eingibt und in der Masse an Suchergebnissen mehr oder weniger zufällig etwas Relevantes findet<sup>2</sup>, sollte ein Nutzer der DDB in einem fachkundig zusammengestellten Bestand bessere Retrievalmöglichkeiten und Rechercheergebnisse im Hinblick auf Qualität und Relevanz erwarten dürfen als im WWW:

Häufig gestaltet sich die Recherche im Netz – ob beruflich oder privat – zu einem bestimmten Thema schwierig: Populäre Suchmaschinen liefern häufig zu viele, zu undifferenzierte Hinweise – oder zu wenige. [...] Ein großer Vorteil: Die DDB stellt nur Informationen zu sorgfältig und fachkundig aufbereitetem Material zur Verfügung. [...] Die auffindbaren Informationen tragen das Gütesiegel deutscher Kultur- und Wissenschaftseinrichtungen. So gibt die DDB echte Orientierungshilfe in der Informationsflut. (Stiftung Preußischer Kulturbesitz, o.J.)

Andererseits sollen sich Nutzer navigierend durch den Bestand bewegen und die medial unterschiedlichen Inhalte je nach ausgewähltem Interessengebiet zusammenhängend erkunden können:

Dabei bietet das Portal [der DDB] auch Gelegenheiten für Entdeckungen und Inspiration. Nutzerinnen und Nutzern werden sich Zusammenhänge erschließen und überraschende Querverbindungen eröffnen. Wer etwa nach ›Beethoven‹ forscht, erhält sofort Sekundärliteratur,

---

<sup>1</sup> Auch themenorientierte oder problemorientierte Suche genannt, vgl. Lewandowski (2015, S. 70); Guha, McCool und Miller (2003, Kap. 2).

<sup>2</sup> Auf die Faktoren des Rankings von Suchergebnissen in Websuchmaschinen kann hier nicht eingegangen werden. Eine umfassende Abhandlung des Themas bietet das Lehrbuch »Suchmaschinen verstehen« von Lewandowski (2015, S. 89-122).

Noten, Portraits – und braucht nur einen Mausklick, um mehr zu erfahren. (Stiftung Preußischer Kulturbesitz, o.J.)

Diese Querverbindungen sollen realisiert werden durch semantische Vernetzungen, die Inhalte über ihre ursprünglichen Erschließungszusammenhänge hinaus miteinander in Beziehung setzen und eine Verfeinerung der Suche durch Filter ermöglichen:

Moderne Such- und Präsentationswerkzeuge eröffnen den Zugang zu hochwertigen digitalen Kopien und vielerlei Werken. Sie ermöglichen bald auch die semantische Verknüpfung der Suchbegriffe, sowohl in Texten als auch in multimedialen Inhalten wie Bildern und Filmen. Durch die vollständige Vernetzung aller Inhalte kann sowohl die Suche als auch die Präsentation der Ergebnisse nach unterschiedlichen Kriterien erfolgen, etwa nach Thema, Ort, Person oder Zeit. (TVT creative media GmbH im Auftrag des Beauftragten der Bundesregierung für Kultur und Medien, 2012)

Tatsächlich bietet das Portal bis heute aber kaum verlässliche Möglichkeiten zum Navigieren in seinem umfangreichen Bestand. Vorhandene Werkzeuge zum Filtern von Suchergebnismengen funktionieren noch nicht zufriedenstellend, wie Untersuchungen zur Retrievalqualität aufgezeigt haben<sup>3</sup>. Noch immer ist das Freitextsuchfeld das vorrangige Recherchewerkzeug in der DDB<sup>4</sup>.

Die besonderen Herausforderungen, denen übergreifende Portale des BAM-Sektors<sup>5</sup> bei der Aufbereitung der Metadaten für ein nutzerfreundliches Retrieval gegenüberstehen, fassen die Informationsspezialisten der Europeana Foundation in einem Strategiedokument zur Verbesserung der Suche in der Europeana folgendermaßen zusammen:

48 million+ records in some 30 languages, of extremely varied nature and quality, [...] heterogeneity, user diversity are all well-recognised challenges in web search, and it cannot be said that Europeana has always met them successfully. [...] How does one navigate and discover 'culture', at scale? (Hill et al., 2016, S. 2)

In der DDB sind es derzeit um 24 Millionen Datensätze zu Kulturobjekten jeglicher Art, die von 369 Institutionen aus dem BAM-Sektor stammen. Diese Metadaten sind in verschiedenen Erfassungssystemen für verschiedene Zielgruppen entstanden und in verschiedenen Formaten exportiert worden. Auf ihrem Weg aus den Erfassungssystemen in die DDB wurden die Metadaten teilweise mehrfach umgewandelt, bereinigt und angereichert, bis sie schließlich in standardisierten Austauschformaten bei der DDB abgeliefert werden konnten<sup>6</sup>. Die verschiedenen Institutionen benutzen unterschiedliche Vokabulare und Verfahren zur Objektbeschreibung und Inhaltserschließung. Viele indexieren nur mit freien Schlagwörtern oder verwenden hauseigene

---

<sup>3</sup> Beispielsweise ist das Filtern von Suchergebnissen in der DDB unbefriedigend und mit Informationsverlust verbunden, vgl. Lindenthal (4. Februar 2016, S. 12–13).

<sup>4</sup> Seb Chan, Direktor Digitale & Neue Technologien am Cooper Hewitt, Smithsonian Design Museum, brachte dieses Problem bei der EuropeanaTech-Konferenz 2015 so auf den Punkt: »Wir haben keine bessere allgemeine Schnittstelle entwickelt als die Suche, um umfangreiche Sammlungen zu präsentieren, wissen aber, dass Suchen nicht der beste Weg ist, etwas zu tun. Es mag der beste Weg sein, wenn du weißt, was du suchst, aber für alles andere ist es schrecklich.« Pekel (2015a).

<sup>5</sup> Im Englischen mit dem Akronym GLAM für Galleries, Libraries, Archives, Museums bezeichnet, das im Rahmen der europäischen Projekte zur Digitalisierung und Sichtbarmachung des Kulturerbes als Oberbegriff für alle darin beteiligten Kulturinstitutionen und Gedächtnisorganisationen verwendet wird.

<sup>6</sup> vgl. Schulze (4. Juni 2018, Folien 4 und 16). Die zitierten Zahlen beziehen sich auf den Stand von Juni 2018.

Vokabulare, die als semantische Anknüpfungspunkte in spartenübergreifenden Portalen nicht geeignet sind. Auch innerhalb einer Sparte sind die Metadaten nicht homogen, weil die Datenelemente der Austauschformate unterschiedlich interpretiert werden und die Erschließungsverfahren voneinander abweichen.

Wie lässt sich bei der unvermeidbaren Heterogenität der Daten das Navigieren und Entdecken im Portal der DDB unterstützen? Welche Voraussetzungen müssen die Metadaten für diese Funktionen erfüllen?

Eine Untersuchung, inwiefern die Metadaten der DDB für eine solche Suchunterstützung geeignet sind, liegt bisher nicht vor.

## **1.2 Forschungsfrage und Untersuchungsziel**

Die übergeordnete Forschungsfrage der vorliegenden Arbeit lautet: Unter welchen Voraussetzungen können die objektbeschreibenden Metadaten der Deutschen Digitalen Bibliothek als navigierbare Elemente verwendet werden, um die Portalnutzer effektiv und auch unabhängig von einer Freitextsuche zu einem möglichst verlässlichen Retrievalergebnis zu führen?

Um dies zu beantworten, werden auf der Grundlage möglicher Suchszenarien die Anforderungen an die Metadaten in einem thematisch, medial und fachlich übergreifenden Portal erarbeitet und mit den Metadaten in den Formaten für die Datenlieferung an die DDB verglichen. Die Ergebnisse dieser Untersuchung sollen eine Grundlage bilden, um Empfehlungen für die Entwicklung eines Verfahrens erarbeiten zu können, in dem die heterogenen Metadaten semantisch soweit vereinheitlicht werden, dass sie als Navigationselemente in der DDB genutzt werden können. Dafür wird folgenden Fragen nachgegangen:

1. Was sind momentan die größten Hindernisse auf dem Weg zu zufriedenstellenden Retrievalergebnissen in der DDB?
2. Wie unterstützen andere Webanwendungen mit umfangreichen und heterogenen Datenbeständen den Suchprozess durch navigierbare Informationsfilter?
3. Welche Informationsfilter sollte die DDB für eine navigierbare Suchunterstützung anbieten, um die Retrievalsituation zu verbessern? Welche Voraussetzungen müssen die objektbeschreibenden Metadaten der DDB für diese Funktionen erfüllen?
4. Welche kontrollierten Vokabulare eignen sich als gemeinsame Bezugspunkte, um bedeutungsgleiche Elementwerte aus den verschiedenen Metadatenformaten zukünftig zusammenführen zu können?
5. Inwiefern eignen sich die Elementwerte der verschiedenen Lieferformate für die DDB tatsächlich für eine navigationsbasierte Suchunterstützung mittels der vorgeschlagenen Informationsfilter?
6. Was wären in Anbetracht des Untersuchungsbefundes die nächsten Schritte auf dem Weg zu einem Verfahren, in dem die Elementwerte über gemeinsame Vokabulare

formatübergreifend vereinheitlicht und als Navigationselemente in der DDB genutzt werden können?

### **1.3 Vorgehensweise und thematische Eingrenzung**

Die Anforderungen an die Metadaten im Hinblick auf navigierbare Informationsfilter im DDB-Portal werden im Rahmen einer Voruntersuchung erarbeitet, deren Ergebnisse in Kapitel 2 vorgestellt werden. Auf der Grundlage von Fachliteratur zum Information-Retrieval werden mögliche Anforderungen und Interaktionen der Benutzer im Suchprozess zusammengefasst und anhand eines fiktiven Suchbeispiels die größten Hindernisse für zufriedenstellende Suchergebnisse im DDB-Portal vorgestellt. Welche Navigationsstrukturen sich aus Nutzersicht in Webanwendungen mit umfangreichen und heterogenen Datenbeständen bewährt haben, fasst der folgende Abschnitt (2.2) zusammen. Dazu werden historische und aktuelle Ansätze verschiedenartiger Webanwendungen kursorisch diskutiert und einzelne Aspekte des Navigierens anhand von Bildschirmfotos ausgewählter Benutzerschnittstellen illustriert. Ausgehend von den vorangegangenen Analysen werden neue navigierbare Informationsfilter zur Unterstützung von informationsorientierten Suchen in der DDB definiert, die zur Verbesserung der Retrievalsituation im Portal beitragen könnten. Aufgrund dieser Vorarbeiten werden die Voraussetzungen an die objektbeschreibenden Metadaten der DDB benannt (siehe Abschnitt 2.3). Die Kriterien zur Auswahl der kontrollierten Vokabulare als Navigationsinstrumente innerhalb der neu definierten Informationsfilter der DDB werden aus der Fachliteratur zum Information-Retrieval und den Standards der Dokumentation und Wissensorganisation sowie Publikationen von Projekten zur Zusammenführung und Vernetzung des digitalen Kulturerbes gewonnen (siehe Abschnitt 2.4).

Den Kern der Arbeit bildet eine Analyse der objektbeschreibenden Metadaten in den verschiedenen Lieferformaten im Hinblick auf ihre Eignung für eine navigationsbasierte Suchunterstützung im Sinne der hier neu vorgeschlagenen Informationsfilter für die DDB. Die Untersuchungsmethode wird in Kapitel 3 erläutert. Zur Bestimmung des Untersuchungskorpus wird eine randomisierte Stichprobe aus der Grundgesamtheit aller Datensätze in der DDB erhoben und deren Aussagekraft abgeschätzt. Ausgangspunkt für die weitere Untersuchung bildet eine Konkordanz der Metadatenelemente aus den verschiedenen Lieferformaten zu den vorgeschlagenen Informationsfiltern der DDB, die anhand eines Abgleiches mit den Spezifikationen und Anwendungsleitfäden der Metadatenschemata und Anwendungsprofile erstellt wird. Auf dieser Grundlage werden die Werte aus den Elementen und Attributen der Datensätze des Untersuchungskorpus abgefragt und für eine Datenauswertung strukturiert aufbereitet. Die so gewonnenen Forschungsdaten werden daraufhin mittels quantitativer und qualitativer Kriterien untersucht. Dabei soll ermittelt werden,

- ob die Metadatenelemente der Lieferformate von den Datengebern der DDB standardkonform angewendet werden und somit verlässlich für die entsprechenden Informationsfilter ausgewertet werden können,

- ob kontrollierte Werte für die Metadatenelemente vorliegen und wenn das der Fall ist, welche kontrollierten Vokabulare verwendet werden, die als Instrumente für navigierbare Informationsfilter in Frage kommen.

Die Ergebnisse dieser Analyse werden geordnet nach Metadatenformat und jeweils untergliedert nach den vorgeschlagenen Filtern in Kapitel 4 präsentiert. Im anschließenden Kapitel 5 werden zunächst die Analyseergebnisse formatübergreifend diskutiert und ein zusammenfassender Befund zur Beantwortung der übergeordneten Forschungsfrage dieser Arbeit formuliert. Davon ausgehend werden die ersten Schritte auf dem Weg zu einem Verfahren aufgezeigt, wie die Metadaten für ein verbessertes Retrieval aufbereitet werden könnten. Abschließend werden Desiderate benannt, die weitere Forschungen anregen können.

In dieser Arbeit werden folgende Themen nicht berücksichtigt: Das Problem der Multilingualität beim Retrieval wird ausgeschlossen, weil die Indexate in der DDB überwiegend in deutscher Sprache vorliegen. Die erheblichen Auswirkungen unterschiedlicher Erschließungspraktiken einzelner Datengeber sowie fehlender objektbeschreibender Metadaten auf die Qualität der Suchergebnisse können im Rahmen dieser Arbeit nicht untersucht werden und erfordern eine eigene Analyse. Die Freitextsuche und die Funktion der Autovervollständigung gehören nicht zum Untersuchungsgegenstand, werden bei der Beschreibung der Ausgangssituation aber einbezogen, um einen Eindruck von den Suchfunktionen in der DDB zu geben. Individualbegriffe (named entities) werden nicht beachtet, weil der Fokus dieser Arbeit auf informationsorientierten beziehungsweise themenorientierten Suchen liegt. Auf Verfahren der automatischen Indexierung kann aufgrund des begrenzten Umfangs dieser Arbeit nur im Ausblick verwiesen werden.

## 1.4 Bezeichnungen und Definitionen

In dieser Arbeit werden die folgenden Bezeichnungen in der hier definierten Bedeutung verwendet:

**Indexterm:** Ein Begriff, repräsentiert durch einen Uniform Resource Identifier (URI) oder eine Bezeichnung, der für die Aufnahme in den Suchindex eines Informationssystems vorgesehen ist. In den Formaten für die Datenlieferung an die DDB ist der Indexterm der Wert eines bestimmten Metadatenelementes.

**Informationsfilter:** Alle navigierbaren Elemente einer Benutzerschnittstelle eines Informationssystems, die Benutzer durch das Auswählen von Hyperlinks zu vordefinierten Ergebnismengen führen. Im Folgenden werden die Bezeichnungen »navigierbarer Informationsfilter« und »Navigationselement« synonym gebraucht.

**Informationsorientierte Suche:** Suche nach Objekten zu einem bestimmten Thema. Im Gegensatz zur Known-Item-Search wird das Informationsbedürfnis des Nutzers dabei nicht durch ein einziges Objekt, sondern durch zusammenhängend auffindbare Objekte befriedigt.

**Kategorien:** Allgemeinbegriffe, die Informationsobjekte unter thematischen oder pragmatischen Gesichtspunkten versammeln und gruppieren und meist in einem Kategoriensystem

präsentiert werden. Der Begriff »Kategorie« bezieht sich in dieser Arbeit nicht auf den philosophisch fundierten Begriff.

**Kontrollierte Vokabulare:** Begriffssysteme, die verbindliche Terme für die Indexierung festlegen und der terminologischen und/oder begrifflichen Kontrolle dienen. Typen von kontrollierten Vokabularen sind Klassifikationen, Thesauri, Normdateien u. a. m.

**Metadatenformat:** Die Abbildung von Instanzdaten in einem bestimmten Metadatenformat.

**Metadatenformat:** Die Spezifikation eines Metadatenformats, in der alle für die Formatierung der Instanzdaten zulässigen Elemente und Attribute definiert sind.

**Objekte:** In einem Informationssystem auffindbare Nachweise von schriftlichen, auditiven, visuellen, audiovisuellen und dreidimensionalen Gegenständen. Im spezifischen Kontext der DDB wird die Bezeichnung »DDB-Objekt« und im allgemeinen Sinne die Bezeichnung »Informationsobjekt« verwendet.

## 1.5 Typografische Gestaltung, orthografische und sprachliche Konventionen

Da der Text auch elektronisch vorliegt, wurde eine serifenlose Schrift gewählt: Standard Calibri 11 Punkt. Für Titel, Zitate und Hervorhebungen werden umgekehrt französische Anführungszeichen, Chevrons, mit den Spitzen nach innen »« verwendet. Diese Anführungszeichen wurden gewählt, um den Text bei häufigem Anführen von Index- und Suchtermen leichter lesbar zu machen. Bezeichnungen für Indexterme und Suchterme stehen in halben Chevrons »«. Individualnamen werden nicht gekennzeichnet.

Metadatenelemente und Attribute sind durch die Schrift Standard Courier New 9 Punkt gekennzeichnet. Bezeichnungen für Metadatenelemente stehen ohne vorangestellten Namensraum in spitzen Winkelklammern <Elemente>. Bezeichnungen für Attribute sind ohne Winkelklammern dargestellt Attribute.

Die Rechtschreibung richtet sich nach Duden »Die neue deutsche Rechtschreibung« in der 27. Auflage. Es wird einheitlich das generische Maskulinum für männliche und weibliche Personen und für Personengruppen verwendet. Datumsangaben sind nach der Norm ISO 8601 gestaltet.

## 2 Voruntersuchung

### 2.1 Suche in der Deutschen Digitalen Bibliothek

Dieses Kapitel fasst mögliche Anforderungen und Interaktionen der Nutzer im Suchprozess zusammen, sofern sie für diese Untersuchung relevant sind. Der Überblick stützt sich auf generische Benutzeranforderungen und einschlägige Literatur aus dem klassischen Information-Retrieval. An einem fiktiven Beispiel werden einige für die DDB spezifizierte Suchfunktionen vorgestellt und die Suchmöglichkeiten im DDB-Portal vergleichend durchgespielt. Als Ergebnis aus den fiktiven Sucherfahrungen werden abschließend die größten Hindernisse für zufriedenstellende Suchergebnisse zusammengefasst.

#### 2.1.1 Allgemeine Benutzeranforderungen und Suchverhalten

Informationsexperten der Europeana Foundation kamen bei der Auswertung einer Personas-Studie für das Europeana-Portal zu dem Schluss, dass kulturinteressierte Nutzer zwei unterschiedliche Anforderungen an die Suche stellen: Einerseits wollen sie einfach und schnell finden können, was sie suchen, andererseits möchten sie den Bestand stöbernd erkunden und dabei Unvorhergesehenes entdecken<sup>7</sup>. Im IFLA<sup>8</sup>-Bibliotheksreferenzmodell (Library Reference Model – LRM) heißen diese Anforderungen »Find« und »Explore«: Beim Finden geht es darum, dass ein Nutzer effektiv zu einem Suchergebnis oder Ergebnis-Set gelangt, das seinem Informationsbedürfnis entspricht. Beim Explorieren geht es darum, dass ein Nutzer Zusammenhänge zwischen Ressourcen entdecken und sich erschließen kann, nach denen er nicht aktiv gesucht hatte. Das Informationssystem unterstützt das Entdecken, indem Beziehungen zwischen Ressourcen explizit gemacht und Funktionen zum Navigieren angeboten werden<sup>9</sup>.

Das Finden und das Explorieren sind in der Regel interaktive Prozesse, in denen ein Nutzer je nach Suchinteresse und Suchunterstützung entweder aktiv sucht oder geführt navigiert. Suchen meint das Formulieren einer Suchanfrage durch die aktive Eingabe eines Suchausdruckes in Suchfelder. Navigieren meint das Auswählen von Hyperlinks, die als Suchunterstützung angeboten werden. Mit einer eingabebasierten Suche erzeugt der Nutzer also selbst eine Ergebnismenge, während eine auswahlbasierte Suche ihn mit Hilfe von Navigationselementen zu vordefinierten Ergebnismengen führt. Navigationselemente sind Retrievalvorschläge, die:

- das Wandern von Ressource zu Ressource über verlinkte Pfade möglich machen
- als Filterwerte zur Verfeinerung von Suchergebnissen dienen können
- zu thematisch verwandten Ergebnismengen führen können
- ein Blättern im Bestand ohne genau bestimmtes Informationsziel unterstützen können

---

<sup>7</sup> vgl. Hill et al. (2016, S. 4).

<sup>8</sup> International Federation of Library Associations and Institutions

<sup>9</sup> vgl. Riva, Le Boeuf und Zumer (2017, S. 16).



- eine weiterführende Suche über verlinkte Schlagwörter ermöglichen<sup>10</sup>.

Ein solches Navigationsangebot setzt voraus, dass die Inhalte des Informationssystems durch gemeinsame kontrollierte Vokabulare erschlossen und semantisch verknüpft sind, wie auch die DDB in ihrer Strategie bis zum Jahr 2020 feststellt<sup>11</sup>. Im IFLA-Bibliotheksmodell sind Anwendungsfälle für die Benutzeranforderungen beschrieben, die u. a. das Navigieren in semantischen Relationen illustrieren<sup>12</sup>. Für »Explore« wird die Möglichkeit in paradigmatischen Hierarchierelationen zu blättern besonders hervorgehoben: Von allgemeinen zu speziellen Begriffen navigierend soll sich ein Nutzer einen strukturierten Überblick über ein Themengebiet verschaffen können.

### 2.1.2 Suchfunktionen in der Anforderungsanalyse der DDB

Unter »Suchen« und »Explorativer Zugang« beschreibt die initiale Anforderungsanalyse für die DDB die angestrebten Suchfunktionen des Portals<sup>13</sup>. Unter »Suchen« fallen das Freitextsuchfeld für initiale und einfache Suchanfragen und die erweiterte Suche für konkretere und zielgerichtete Suchen. Alle Freitextsuchfelder sollen über eine Autocomplete-Funktion verfügen, die automatisch Vorschläge zur Vervollständigung der Sucheingabe anbietet. Diese Funktion soll Synonyme und verwandte Begriffe sowie normierte Schlagwörter der DDB einbeziehen und Disambiguierungsvorschläge für mehrdeutige Begriffe bereitstellen. Für »weniger beziehungsweise gar nicht zielgerichtete Suchen« sieht die Anforderungsanalyse verschiedene explorative Zugänge vor, unter anderen<sup>14</sup>:

- Navigieren über vordefinierte Themengebiete
- Filterfunktion zur nachträglichen Verfeinerung einer Suchanfrage
- Verweise auf weiterführende Informationen in den Detailansichten von Objekten.

### 2.1.3 Suchfunktionen im DDB-Portal an einem Beispiel

Um einen Eindruck von der Suchunterstützung sowie den Hürden bei der Suche in der DDB zu geben, werden im Folgenden die angebotenen Suchfunktionen an vier fiktiven Szenarien durchgespielt<sup>15</sup>. Die Suchen wurden im August 2018 wie beschrieben durchgeführt und die Ergebnisse faktengetreu dokumentiert.

Die Szenarien basieren auf folgendem Suchbeispiel: Ein geübter Internetnutzer sucht Informationen und Materialien zur Einführung in die künstlerische Malerei mit Ölfarben. Ihn

<sup>10</sup> vgl. Hearst (2009, Kapitel 8); Villa et al. (2013, S. 19-22)

<sup>11</sup> siehe Stiftung Preußischer Kulturbesitz (2016, S. 34).

<sup>12</sup> siehe Riva, Le Boeuf und Zumer (2017, S. 97-99).

<sup>13</sup> siehe Fraunhofer IAIS (21. Mai 2010, S. 57-69).

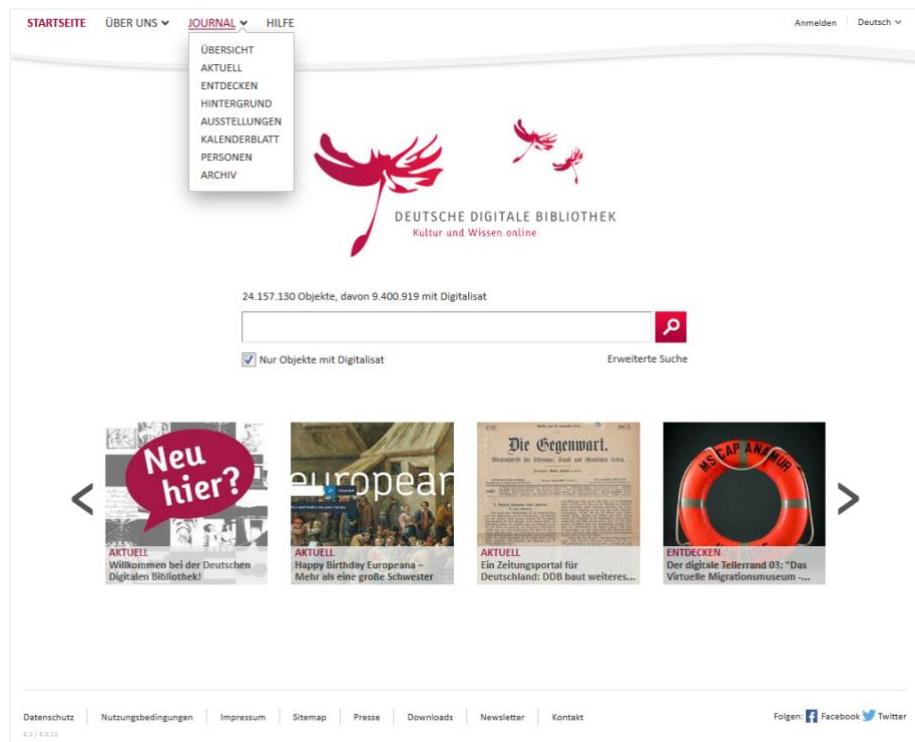
<sup>14</sup> vgl. Fraunhofer IAIS (21. Mai 2010, S. 60-64).

<sup>15</sup> Für die erweiterte Suche wird kein Szenario vorgestellt, weil sie in der DDB kaum benutzt wird. Im Oktober 2018 wurde bei 14,3 Prozent aller Besuche eine Freitextsuche ausgelöst (13.677 von 95.590 Besuchen), aber nur bei 0,6 Prozent eine erweiterte Suche (536 von 95.590 Besuchen). Diese Zahlen wurden am 15.11.2018 mit der Webanalytik-Plattform Matomo Version 3.6.1 erhoben.

interessiert vor allem Literatur zur Malerei mit dem Schwerpunkt auf der Technik der Ölmalerei, und er möchte auch Beispiele für Ölgemälde finden. Er hat also eine grobe Zielvorstellung, aber noch keine genaue Suchstrategie. Daher möchte er sich zunächst einen Überblick über das Angebot in der DDB zu seinem Themenschwerpunkt verschaffen und im späteren Suchverlauf seine Recherche auf einzelne Aspekte konzentrieren.

Aus seinen Internet-Recherchen kennt der Nutzer die Autovervollständigungsfunktion zur Anfrageformulierung und die Filterfunktionen zur Verfeinerung von Ergebnismengen. Er will diese Funktionen nutzen, denn er verspricht sich von vordefinierten Ergebnismengen bessere Trefferquoten und größere Genauigkeit als bei Freitextsuchen.

Die Startseite bietet auf den ersten Blick keine auswahlbasierte Suchunterstützung: Abgesehen von dem Hinweis, wie viele Objekte in der DDB insgesamt vorhanden sind, erhält der Nutzer keinerlei Auskunft darüber, was er im Portal finden kann:



**Abbildung 2.1: Startseite des DDB-Portals<sup>16</sup>**

Um sich den gewünschten Überblick zu seinem Thema zu verschaffen, steht der Nutzer daher vor der Herausforderung, sich durch das Eingeben eines Suchausdruckes in das Freitextsuchfeld selbst eine Ergebnismenge erzeugen zu lassen. Wie soll er seine initiale Suchanfrage formulieren, um die Sprache des Informationssystems zu treffen? Soll er:

- nach ›Malerei‹ suchen und auf Verbindungen zu passenden Gemälden setzen,
- nach ›Gemälde‹ suchen und auf verknüpfte Informationen zum Thema Malerei hoffen,

<sup>16</sup> Quelle: Deutsche Digitale Bibliothek, URL: <http://www.deutsche-digitale-bibliothek.de> (zitiert am: 30.11.2018)

c) spezifisch nach ›Ölmalerei‹ oder nach ›Ölgemälde‹ fragen?

Im fachlich und medial übergreifenden Bestand der DDB sind die Bezeichnungen dieser möglichen Suchterme nicht eindeutig: ›Malerei‹ kann (1) die künstlerische Gattung oder (2) die Maltechnik oder (3) das realisierte Kunstwerk meinen. Während sich ›Gemälde‹ eindeutig auf das Kunstwerk bezieht, kann der spezifischere Ausdruck ›Ölmalerei‹ sowohl die Technik als auch das realisierte Werk meinen. Für den Sucheinstieg entscheidet sich der Nutzer für Variante (a) und wählt den umfassenderen generischen Einfachbegriff ›Malerei‹ in der Annahme, dass er damit eine möglichst vollständige oder zumindest umfassende Auswahl relevanter Treffer erhält und ihm im Suchverlauf Vorschläge zum Einschränken der Suchergebnisse gemacht werden.

### Szenario 1: Sucheinstieg über die Freitextsuche nach ›malerei‹

Die Suchanfrage nach ›malerei‹ liefert etwa 65.000 Ergebnisse<sup>17</sup>. Der Nutzer blättert vielleicht die ersten beiden Ergebnisseiten durch<sup>18</sup>: Er findet überwiegend Druckwerke zum Thema Malerei, darunter 17 Kapitel einzelner Bände einer Schriftenreihe, diverse Bildbände ohne Digitalisat und acht Fotografien mit dem Titel »Sowjetische Malerei«, aber nichts Interessantes zu seinem Suchinteresse. Er findet auch keinen Anhaltspunkt, wie er eine Auswahl relevanter Objekte identifizieren könnte:

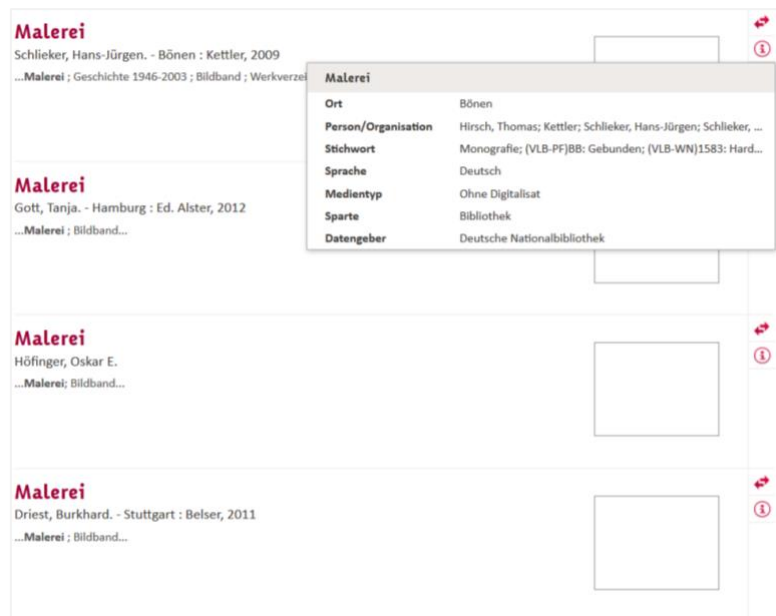


Abbildung 2.2: Ergebnisse für Suchanfrage nach ›malerei‹ in der DDB<sup>19</sup>

<sup>17</sup> Bei allen Suchanfragen wurde der voreingestellte Filter »Nur Objekte mit Digitalisat« unter dem Freitextsuchfeld deaktiviert. Die Zahlen zu den Ergebnismengen sind grob gerundet.

<sup>18</sup> Für die Suchergebnisanzeige wurde folgende Voreinstellung benutzt: Listenansicht, Sortierung der Ergebnisse nach Relevanz, 20 Ergebnisse pro Seite. Unterschiedliche Untersuchungen haben ergeben, dass etwa 80 Prozent der Nutzer sich nur die ersten zehn Treffer in der Ergebnisliste, also in der Regel die erste Ergebnisseite, ansehen, vgl. Lewandowski (2005, S. 38).

<sup>19</sup> Quelle: Deutsche Digitale Bibliothek, URL: <http://www.deutsche-digitale-bibliothek.de> (zitiert am: 23.08.2018)

Im Stichwortfilter<sup>20</sup> der DDB hofft der Nutzer, passende Vorschläge zur Verfeinerung der Suche zu finden. Da ihn speziell »Ölmalerei« interessiert, schränkt er die Stichwortauswahl auf Werte mit dem Bestandteil »öl« ein:

The screenshot shows the DDB search interface. On the left, there is a sidebar with filters: 'Ergebnisse filtern', 'Zeit', 'Ort', 'Person/Organisation', 'Stichwort', 'Sprache', 'Rechtsstatus', 'Verwendbarkeit', 'Medientyp', 'Sparte', and 'Datengeber'. The 'Stichwort' filter is active, and the search term 'öl' is entered. The main area shows 64,624 results, with a tab for 'OBJEKTE' selected. Below the search bar, the results are displayed as a list of terms with their counts. The terms are: 'Ölgemälde' (54), 'Gemälde (Ölgemälde im Holzrahmen)' (16), 'Projekt: Ölgemälde' (9), 'Ölbild' (9), 'Gemälde (Ölgemälde)' (7), 'Ölstudie' (7), 'Ölmalerei' (5), 'Gethsemane, Ölberg (Matthäus 26:36-56; Markus 14:32-52; Lukas 22:39-53; Johannes 18:1-12)' (3), 'Bäume: Ölbaum' (2), and 'Gemälde (Ölgemälde ohne Rahmen)' (2). The results are sorted by frequency ('Nach Häufigkeit geordnet').

**Abbildung 2.3: Werte mit Bestandteil »öl« im Stichwortfilter der DDB<sup>21</sup>**

Die Liste enthält eine Fülle synonyme Bezeichnungen<sup>22</sup>. Für »Ölmalerei« gibt es dagegen nur fünf Treffer; 54 Einträge sind für »Ölgemälde« ausgewiesen<sup>23</sup>, 53 davon stammen aber von ein und demselben Datengeber und zeigen fast ausnahmslos Porträts, die dem Interesse des Nutzers nicht entsprechen.

## Szenario 2: Sucheinstieg mit Hilfe eines Suchvorschlages

Beim Eintippen von »malerei« in das Freitextsuchfeld wird der Nutzer durch die Funktion der Autovervollständigung auf den Vorschlag »malerei gemälde« aufmerksam. Aktuell (23.08.2018) enthält die Auswahlliste zu »malerei« erkennbar keine kontrollierten Terme, Synonymverweisungen oder Disambiguierungen, sondern ausschließlich Freitextphrasen:

<sup>20</sup> Der Stichwortfilter müsste eigentlich Schlagwortfilter heißen, weil nur Terme aus Indexteilen in diesen Filter eingehen, vgl. Kapitel 3.2 zur Auswahl der Elemente aus den Metadatenschemata.

<sup>21</sup> Quelle: Deutsche Digitale Bibliothek, URL: <http://www.deutsche-digitale-bibliothek.de> (zitiert am: 23.08.2018)

<sup>22</sup> Stichwörter mit dem Bestandteil »öl« in Szenario 1: »Ölgemälde« (54) ; »Gemälde (Ölgemälde im Holzrahmen)« (16) ; »Projekt: Ölgemälde« (9) ; »Ölbild« (9) ; »Gemälde (Ölgemälde)« (7) ; »Ölstudie« (7) ; »Ölmalerei« (5) ; »Gethsemane, Ölberg (Matthäus 26:36-56; Markus 14:32-52; Lukas 22:39-53; Johannes 18:1-12)« (3) ; »Bäume: Ölbaum« (2) ; »Gemälde (Ölgemälde ohne Rahmen)« (2) ; »Gemälde (Ölstudie)« (2) ; »Ölskizze« (2) ; »Bruno, der Gründer des Kartäuserordens; mögliche Attribute: Kreuz, Finger an den Lippen, Lilie, Mitra zu seinen Füßen, Ölweig, Lichtstrahl, Totenschädel, (sieben) Stern(e)« (1) ; »Christus und seine Apostel auf dem Ölberg« (1) ; »Gemälde (Öl auf Leinen)« (1) ; »Gemälde (Ölgemälde im Metallrahmen)« (1).

<sup>23</sup> Das Stichwort »Ölgemälde« führt allerdings zu 58 und nicht zu 54 Ergebnissen. Die Anzahl ist hier im Filter falsch ausgewiesen.



**Abbildung 2.4: Suchvorschläge für ›malerei‹ im Freitextsuchfeld der DDB<sup>24</sup>**

Der Nutzer übernimmt den Suchvorschlag ›malerei gemälde‹<sup>25</sup> in der Erwartung mit diesem komplexen Ausdruck zu einer passenderen Ergebnisauswahl geführt zu werden. Die noch immer große Ergebnismenge (11.800), die auf den ersten Seiten keine passenden Treffer bietet, möchte der Nutzer wieder mit Unterstützung des Stichwortfilters einschränken. Die Vorschläge ›Malerei‹ (7.000) und ›Gemälde‹ (5.500) führen erneut zu großen Treffermengen. Die 1.900 Treffer für das Stichwort ›Malerei/Gemälde‹ stammen ausschließlich von einem einzigen Datengeber. Für ›Ölmalerei‹ gibt es nur einen Treffer und für ›Ölgemälde‹ lediglich 52 Ergebnisse, die logischerweise eine Teilmenge der Filterergebnisse aus Szenario 1 repräsentieren<sup>26</sup>.

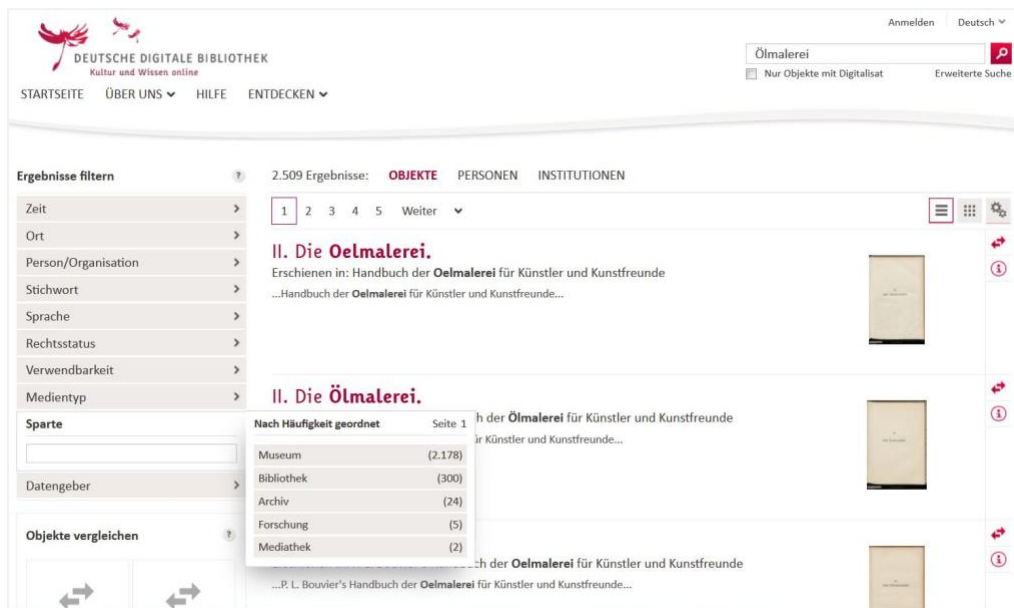
### **Szenario 3: Reformulieren der Suchanfrage**

Da die Ergebnismengen der vorangegangenen Sessions zu groß und sich nicht sinnvoll einschränken ließen, spezifiziert der Nutzer seine Anfrage durch den differenzierten Suchausdruck ›ölmalerei‹ und erhält circa 2.500 Treffer, von denen ihn schon die ersten Ergebnisse interessieren. Die Mehrzahl der gefundenen Objekte (2.200) stammt aus Museen:

<sup>24</sup> Quelle: Deutsche Digitale Bibliothek, URL: <http://www.deutsche-digitale-bibliothek.de> (zitiert am: 23.08.2018)

<sup>25</sup> In der Freitextsuche der DDB ist der boolesche Operator AND für die Verknüpfung von Suchtermen voreingestellt.

<sup>26</sup> Stichwörter mit dem Bestandteil ›öl‹ in Szenario 2: ›Ölgemälde‹ (52) ; ›Gemälde (Ölgemälde im Holzrahmen)‹ (16) ; ›Ölbild‹ (9) ; ›Gemälde (Ölgemälde)‹ (7) ; ›Gemälde (Ölgemälde ohne Rahmen)‹ (2) ; ›Gemälde (Ölstudie)‹ (2) ; ›Ölskizze‹ (2) ; ›Gemälde (Öl auf Leinen)‹ (1) ; ›Gemälde (Ölgemälde im Metallrahmen)‹ (1) ; ›Ölmalerei‹ (1) ; ›Ölstudie‹ (1)



**Abbildung 2.5: Ergebnisse für Suchanfrage nach ›Ölmalerei‹ in der DDB<sup>27</sup>**

Im Stichwortfilter finden sich aber nur 86 Einträge für ›Ölmalerei‹, von denen wiederum 83 aus Bibliotheken stammen<sup>28</sup>.

#### **Szenario 4: Suchunterstützung auf der Detailseite**

Bei der Recherche ist dem Nutzer ein Ölgemälde mit dem Titel »Die Malerei« besonders aufgefallen und er sieht sich das Objekt im Detail an, auch um möglicherweise weitere Suchunterstützung zu bekommen (siehe Abbildung 2.6). Auf der Detailseite des Objektes ist der Name des Malers mit dem entsprechenden Personendatensatz der DDB verlinkt, welcher wiederum mit weiteren Werken des Künstlers verknüpft ist, die in der DDB vorhanden sind und die der Nutzer bei Interesse durchblättern kann<sup>29</sup>. Den Nutzer spricht jedoch vor allem das Sujet an, die Malerei selbst als Thema, künstlerisch gestaltet durch ein »Bild im Bild«. In der Detailansicht findet er aber keine verknüpften Einträge oder andere suchunterstützenden Elemente, die ihn bei der weiteren Suche nach »Malerei als Thema« leiten könnten. Es gibt keinen nach Inhalt differenzierten Zugang zu Schlagwörtern, die das Thema oder das auf visuellen Objekten Abgebildete beschreiben.

<sup>27</sup> Quelle: Deutsche Digitale Bibliothek, URL: <http://www.deutsche-digitale-bibliothek.de> (zitiert am: 23.08.2018)

<sup>28</sup> Jeweils ein Objekt zum Schlagwort ›Ölmalerei‹ stammt aus den Sparten Archiv, Forschung und Museum.

<sup>29</sup> siehe <http://www.deutsche-digitale-bibliothek.de/entity/174313780>

Datengeber

→ Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg

www.fotomarburg.de/

Objekt beim Datengeber anzeigen

Link auf diese Seite

## Die Malerei

**Objektbezeichnung:** Supraporte

**Material/Technik:** Öl; Leinwand

**Ereignis:** Herstellung

**(wer):** Heinsius, Johann Ernst (1740)

**(wann):** 1762/1764

**Bezug (was):** symbolische Darstellungen, Allegorien und Embleme der Malerei; Pictura; Ripa: Pittura

**Klassifikation:** Malerei (Objektklassifikation)

**Standort:** Schloss Heidecksburg, Grüner Saal, Rudolstadt

**Rechtsstatus:** © Rechte vorbehalten - Freier Zugang

Objekt beim Datengeber anzeigen

Letzte Aktualisierung: 15.04.2018, 06:56 Uhr

Die Malerei  
Foto: Nehrlich, Rolf W. (1943/1944)  
© Bildarchiv Foto Marburg

**Bilder (3)**

Abbildung 2.6: Detailseite für das Objekt »Die Malerei« in der DDB<sup>30</sup>

### 2.1.4 Zusammenfassung und Schlussfolgerung

Das Suchbeispiel zeigt, dass der Zugang zu gewünschten Informationen in der DDB nicht einfach ist. Ein Hindernis für eine erfolgreiche Informationsbeschaffung sind große Ergebnismengen, die sich nicht zuverlässig auf Relevantes einschränken lassen. Der Stichwortfilter kann bei der Sucheinschränkung kaum helfen: Er versammelt unterschiedslos Schlagwörter aus verschiedenen Metadatenelementen, die

- den Typ oder die Gattung der Kulturobjekte kennzeichnen, wie z. B. »Ölgemälde«,
- den Inhalt beschreiben, z. B. »Ölmalerei« als Thema eines gedruckten Werkes,
- die Objekte grob kategorisieren, z. B. »Malerei« als Werk der bildenden Kunst.

Es gibt keinen navigierbaren Sucheinstieg, anhand dessen sich ein Nutzer einen Überblick über das Angebot der DDB verschaffen kann. Für den Sucheinstieg wie im weiteren Suchverlauf fehlen sinnvolle Vorschläge zur Verbesserung oder Verfeinerung des Suchergebnisses: Weder die Autovervollständigung noch die Filter können bei der Präzisierung informationsorientierter Suchen helfen, weil das Angebot an gemeinsam genutzten, kontrollierten Indextermen in der Freitextsuche und in den Filtern zu gering ist. Im Stichwortfilter finden sich kontrollierte und unkontrollierte

<sup>30</sup> Quelle: Deutsche Digitale Bibliothek, URL: <http://www.deutsche-digitale-bibliothek.de/item/RFMWXNWCLL7WLZ4NC76RIKX7WN25GUPJ> (zuletzt aktualisiert am: 25.04.2018)

Terme, Oberbegriffe neben Unterbegriffen und eine Fülle von Synonymen und Homonymen undifferenziert beieinander<sup>31</sup>.

Darin liegt auch ein Grund für den enormen Informationsverlust, den die Filterwerte mit sich bringen. Die oben vorgestellten Ergebnisse für Vorschläge aus dem Stichwortfilter zeigen deutlich, dass die Vollständigkeit der Terme (Recall) bei diesen Navigationselementen ausgesprochen gering ist. Beispielsweise zeigt der Stichwortfilter in der Ergebnismenge einer Freitextsuche nach ›öl‹ etwa 17.700 Treffer für den Filterwert ›gemälde‹ an, eine Zahl, die im Verhältnis zu 109 Treffern für das Stichwort ›ölgemälde‹ völlig unplausibel ist<sup>32</sup>. Auch bei einer Vereinigung mit den sieben Synonymen, die der Filter für ›ölgemälde‹ anbietet, käme der Nutzer auf nur 146 Ergebnisse<sup>33</sup>. Ein ganz entscheidender Faktor für den Informationsverlust deutet sich hier bereits an: Die Schlagwörter, die im Stichwortfilter angeboten werden, sind eher spezifisch für die datengebende Institution oder eine bestimmte Sparte als für das Objekt. Der Deskriptor ›Ölmalerei‹ zum Beispiel wird fast ausschließlich von Bibliotheken vergeben, während ›Ölgemälde‹ fast nur von einer Handvoll Museen indexiert wird.

Der Informationsverlust geht auch auf die fehlende Auswertung paradigmatischer Hierarchierelationen zwischen Oberbegriffen und Unterbegriffen zurück<sup>34</sup>: Die Menge von circa 60.000 Ergebnissen für das Stichwort ›malerei‹ erhöht sich beispielsweise durch die Vereinigung mit ›ölmalerei‹, ›tafelmalerei‹, ›gemälde‹ und ›ölgemälde‹ auf beinahe 108.000 Treffer<sup>35</sup>.

Weitere Gründe für den schwachen Recall können einerseits auch eine zu grobe oder fehlende Erschließung durch den Datengeber sein sowie andererseits eine fehlende Auswertung der gelieferten Indexterme durch die DDB. Diese Hindernisse für ein qualitätsvolles Retrieval können allerdings im Rahmen dieser Arbeit nicht systematisch untersucht werden.

Das Venn-Diagramm in Abbildung 2.7 illustriert den Informationsverlust bei der Suche im DDB-Portal. Es zeigt, dass semantisch in die Extension und das Wortumfeld des Begriffes ›Malerei‹ fallende »Stichwörter« nur mit geringen Überschneidungen mit dem Stichwort ›Malerei‹ verbunden sind. Nicht einmal die generischen Unterbegriffe ›Ölmalerei‹, Wandmalerei‹, ›Freskomalerei‹ u. a. sind in die Ergebnismenge von ›Malerei‹ vollständig eingeschlossen.

---

<sup>31</sup> Synonyme sind zum Beispiel ein Grund für Informationsverlust (verminderter Recall), Homonyme sind ein Grund für Informationsballast (verminderte Präzision) beim Retrieval.

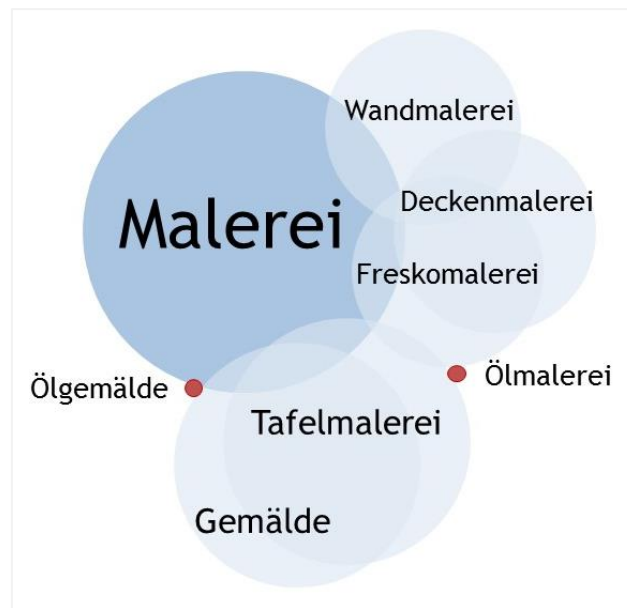
<sup>32</sup> Diese Suche wurde am 24.08.2018 durchgeführt.

<sup>33</sup> Aus: ›ölgemälde‹ (109) OR ›Gemälde (Ölgemälde im Holzrahmen)‹ (16) OR Projekt: ›Ölgemälde‹ (9) OR ›Gemälde (Ölgemälde)‹ (7) OR ›Gemälde (Ölgemälde ohne Rahmen)‹ (2) OR ›Gemälde (Ölgemälde im Metallrahmen)‹ (1) OR ›Gemälde (Ölgemälde im Holzrahmen)‹ (1) OR ›Ölgemälde (gerahmt)‹ = 146 Ergebnisse. Im Stichwortfilter der DDB ist der boolesche Operator OR für die Verknüpfung verschiedener Filterwerte voreingestellt.

<sup>34</sup> Speziell die Hyponym-Hyperonym-Relation oder Abstraktionsbeziehung oder generische Relation, vgl. ISO 25964-1:2011 (Clause 10.2); Stock und Stock (2008, S. 75 ff.)

<sup>35</sup> Diese Vereinigungsmenge wurde am 27.08.2018 mit der Funktion der erweiterten Suche gebildet. Die Suchanfrage »keywords:(malerei) OR keywords:(tafelmalerei) OR keywords:(gemälde) OR keywords:(ölgemälde) OR keywords:(ölmalerei)« ergab 107.654 Ergebnisse.





**Abbildung 2.7: Informationsverlust bei der Suche im DDB-Portal am Beispiel ›Malerei‹**

Dem Diagramm liegt der Suchindex der DDB vom 23.08.2018 zu Grunde. Ausgewertet wurden die Ergebnisse der erweiterten Suche für die acht dargestellten Stichwörter, die jeweils paarweise mit dem booleschen Operator AND verknüpft wurden. Die Kreisflächen sind proportional zur Trefferanzahl dargestellt, die Schnittmengen dagegen nur annähernd proportional

Eine Suchausweitung über Hierarchierelationen wäre im Aggregationsportal der DDB allerdings auch nur sinnvoll möglich, wenn die Kulturobjekte spartenübergreifend mit geeigneten kontrollierten Vokabularen erschlossen wären. Inwiefern dies der Fall ist, soll im Rahmen dieser Arbeit untersucht werden.

## 2.2 Informationsfilter in Webanwendungen

Schon im ursprünglichen Konzept für das World Wide Web (WWW) formulierte sein Erfinder Tim Berners-Lee im Jahr 1989, dass wortbasierte Verfahren allein die Suche in komplexen Informationsräumen nicht ausreichend unterstützen:

Keywords are a common method of accessing data for which one does not have the exact coordinates. The usual problem with keywords, however, is that two people never chose the same keywords. The keywords then become useful only to people who already know the application well. [...] The system must allow any sort of information to be entered. Another person must be able to find the information, sometimes without knowing what he is looking for. [...] Much of the academic research is into the human interface side of browsing through a complex information space. Problems addressed are those of making navigation easy, and avoiding a feeling of being ›lost in hyperspace‹. (Berners-Lee, March 1989, May 1990)

Der Bedarf an Benutzerschnittstellen, die ein einfaches Navigieren durch komplexe Datenbestände ermöglichen, wurde bereits früh erkannt. In den 1990er Jahren kamen daher mit den Webverzeichnissen die ersten hierarchisch organisierten Suchsysteme auf, die einen navigierbaren

Einstieg in die Recherche ermöglichen. Das Navigieren über Kategorien, also über vordefinierte Gruppen von Informationsobjekten, dient bis heute in vielen Webanwendungen als erste Einschränkung zu Beginn der Suche. Mit der Erkenntnis, dass die Navigationswege über die anfänglich monohierarchisch organisierten Kategoriensysteme zu starr und unflexibel sind, begann in der ersten Hälfte der 2000er Jahre die Entwicklung facettierter Suchsysteme<sup>36</sup>. Vorreiter waren mit den Portalen des elektronischen Handels kommerzielle Webanwendungen<sup>37</sup>, die kategorienbasierte Sucheinstiege mit facettenbasierten Filtern zum Einschränken von Suchergebnismengen kombinieren.

Dieses Kapitel gibt einen kursorischen Überblick über erste Ansätze und aktuelle Umsetzungen von navigierbaren Informationsfiltern in verschiedenartigen Webanwendungen mit umfangreichen und heterogenen Datenbeständen. Ein besonderes Augenmerk liegt darauf, wie die Benutzerschnittstellen das Navigieren als Teil des Suchprozesses unterstützen und welche Arten von Informationsfiltern heutigen Nutzerwünschen am meisten entsprechen.

### 2.2.1 Webverzeichnisse

In der Anfangszeit des WWW boten Webverzeichnisse oder Webkataloge über hierarchisch angeordnete Kategorien einen navigierbaren Einstieg zu den Inhalten des WWW<sup>38</sup>. Um ein Blättern vom Allgemeinen zum Speziellen zu unterstützen, wurden thematisch verwandte Websites in einem redaktionellen Prozess händisch selektiert und hierarchisch kategorisiert<sup>39</sup>. Die Nutzung der universellen Webverzeichnisse ging aber mit dem Vormarsch der Websuchmaschinen deutlich zurück, bis viele von ihnen schließlich ganz eingestellt wurden<sup>40</sup>. Der Niedergang der Webverzeichnisse ist neben den hohen personellen Aufwänden und Kosten vor allem darin begründet, dass die Anbieter eigene Kategoriensysteme entwickelten, die mit stetigem Anwachsen zu Inkonsistenzen und kaum vorhersehbaren Navigationsstrukturen führten<sup>41</sup>.

### 2.2.2 Subject Gateways

Während universelle Webverzeichnisse alle Inhalte des WWW indizierten, zielen Subject Gateways<sup>42</sup> auf eine Auswahl als wissenschaftlich relevant eingeschätzter Internetressourcen, um sie für Forschung und Lehre zugänglich zu machen. Als wesentliches Merkmal der Subject Gateways

---

<sup>36</sup> vgl. Hearst (2009, Kapitel 8.6).

<sup>37</sup> Als Beispiel für einen frühen Versuch aus der Wissenschaft zur Entwicklung einer Facettensuche für eine bestandsübergreifende Webanwendung sei hier auf die Machbarkeitsstudie des Projektes »Facet Analytical Theory in Managing Knowledge Structure for Humanities« (FAT-HUM) verwiesen, vgl. Broughton und Slavic (2007).

<sup>38</sup> Das älteste Webverzeichnis, The WWW Virtual Library, erreichbar unter der URL <http://vlib.org/>, wurde 1991 von Tim Berners Lee entwickelt. Die großen universellen Webverzeichnisse wie das Yahoo Directory (1995-2014) und das Open Directory Project (ODP) mit dem alternativen Namen dmoz (1998-2017) sind mittlerweile eingestellt. dmoz ist noch unter der URL <http://dmoztools.net/> als Archivversion zugänglich. Unter dem Namen Curlie wird das Verzeichnis von ODP/dmoz seit 2017 weitergeführt, URL: <http://curlie.org/>.

<sup>39</sup> vgl. Lewandowski (2015, S. 23f.).

<sup>40</sup> vgl. Golem.de (2014); Young (2011)

<sup>41</sup> vgl. Hearst (2009, Kapitel 8.5 und 8.7); Lewandowski (2005, S. 204f.); Koch und Day (1997, S. 5).

<sup>42</sup> auch: Fachportale, virtuelle Fachbibliotheken, Fachinformationsführer

wurde ein browsender Sucheinstieg über eine Sachsystematik gesehen<sup>43</sup>. Im von der Europäischen Kommission geförderten Projekt Renardus (2000-2002) wurde beispielsweise ein Pilot-Service entwickelt, der über eine Benutzeroberfläche einen klassifikationsbasierten Sucheinstieg in die verteilten Datenbestände von zwölf europäischen Fachportalen ermöglichen sollte. Die Dewey Decimal Classification (DDC) wurde als integrierendes Begriffssystem gewählt und über eine »Cross-Browsing-Funktionalität« zum Navigieren angeboten<sup>44</sup>. Auch wenn der Service nicht umgesetzt wurde, bleibt im Zusammenhang mit dem Navigieren über Kategorien eine damals überraschende Erfahrung erwähnenswert: Entgegen der ursprünglichen Annahme zogen Nutzer einen browsenden Sucheinstieg, wenn er denn angeboten wurde, einer Freitextsuche eindeutig vor<sup>45</sup>.

### 2.2.3 Portale des elektronischen Handels

Das »Stöbern in Kategorien« hat sich in Portalen des elektronischen Handels wie alibaba<sup>46</sup>, Amazon<sup>47</sup> oder ebay<sup>48</sup> etabliert. Mit einem Kategoriensystem knüpft eine Website an Nutzergewohnheiten an<sup>49</sup>. Ein Ausschnitt des Kategoriensystems, welches einen ersten Überblick über das Angebot gibt und eine gute Vorauswahl am Anfang der Suche ermöglicht, wird häufig in einem aufklappbaren Menü in der Nähe des Sucheingabefeldes angezeigt. Darüber könnte der Nutzer aus dem fiktiven Suchbeispiel, das in Kapitel 2.1.3 dieser Arbeit beschrieben ist, direkt mit einer ausgewählten Kategorie in die Suche einsteigen, ohne einen Suchausdruck in das Suchfeld eingeben zu müssen. Für die Suche nach Informationen zur künstlerischen Malerei mit Ölfarben wäre zum Beispiel die passende Einstiegskategorie bei ebay »Antiquitäten & Kunst«<sup>50</sup>:

---

<sup>43</sup> vgl. Koch (2000, S. 24f.).

<sup>44</sup> vgl. Koch, Neuroth und Day (2001, S. 17ff.) Eine kurze Zusammenfassung des Verfahrens zur Herstellung semantischer Interoperabilität zwischen der DDC und den Klassifikationen der projektbeteiligten Nationalbibliotheken und Forschungszentren findet sich im Aufsatz von Hubrich (2018, S. 283f.).

<sup>45</sup> vgl. Koch, Golub und Ardö (2006, S. 26).

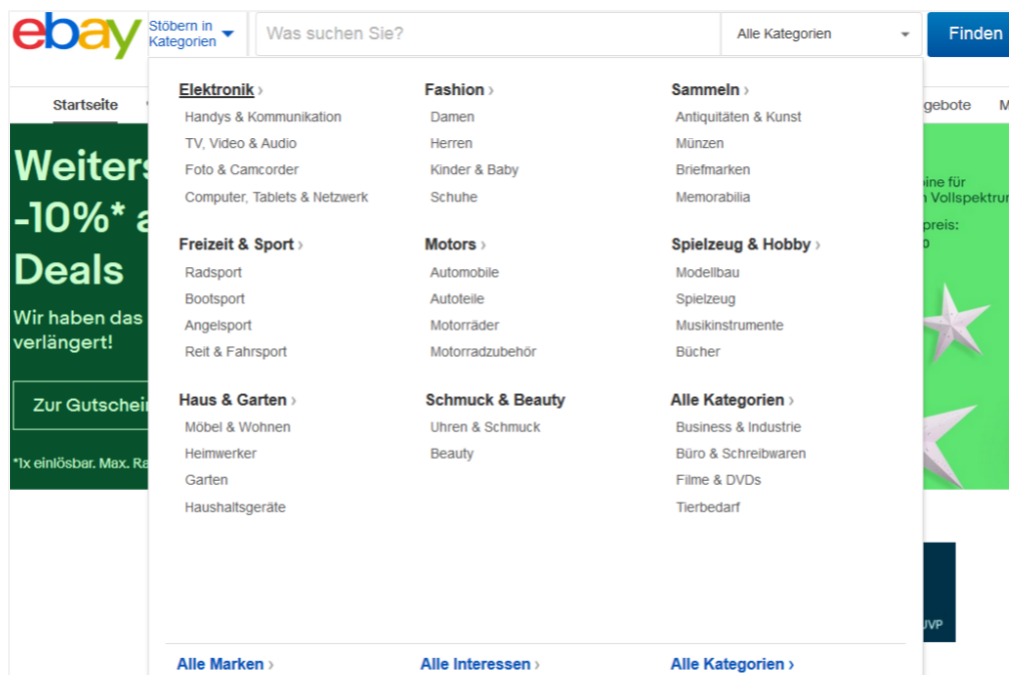
<sup>46</sup> siehe <http://www.alibaba.com/> und <http://www.alibaba.com/Products>

<sup>47</sup> siehe <https://www.amazon.de/> und <https://www.amazon.de/gp/site-directory/>

<sup>48</sup> siehe <https://www.ebay.de/> und <https://www.ebay.de/v/allcategories>

<sup>49</sup> vgl. Hearst (2009, Kapitel 8.1).

<sup>50</sup> Aktuell (01.12.2018) wird der Suchraum über diese Vorauswahl auf circa 4,4 Millionen Artikel eingeschränkt.



**Abbildung 2.8: »Stöbern in Kategorien« bei ebay<sup>51</sup>**

Die Kategoriensysteme der Verkaufskataloge sind im Gegensatz zu denen der Webverzeichnisse überschaubar: Sie haben wenige Hierarchiestufen (i. d. R. nicht mehr als drei oder vier) und eine begrenzte Anzahl an Einstiegskategorien auf der obersten Ebene (alibaba: 12, Amazon: 19, ebay: 9 am 01.12.2018). Das vollständige Kategoriensystem ist über einen Hyperlink, bei ebay »Alle Kategorien« genannt, erreichbar<sup>52</sup>.

Die Kategorien kommen auch in Kombination mit einer Suchanfrage zum Einsatz und werden u. a. zur Präzisierung und Disambiguierung von Suchvorschlägen angeboten. Auf der Suche nach Materialien zum Thema »Ölmalerei« etwa könnte ein ebay-Nutzer bei der Eingabe des Suchausdruckes »Ölmalerei« entscheiden, ob er diesen entweder (a) mit der Kategorie »Kunst« oder (b) mit der Kategorie »Sachbücher« kombinieren möchte (siehe Abbildung 2.9). Der erste Suchvorschlag bietet 60.000 Ergebnisse, die allesamt Ölgemälde sind. Die zweite Auswahl liefert 250 Bücher und einige DVDs über die Technik der Ölmalerei, die dem Suchinteresse des Nutzers entsprechen würden<sup>53</sup>. Darüber hinaus bieten die Verkaufskataloge i. d. R. durch ein Auswahlmenü neben dem Suchfeld die Möglichkeit, eine Kategorie mit einer Suchanfrage zu kombinieren.

<sup>51</sup> Quelle: ebay, URL: <http://www.ebay.de> (zitiert am: 01.12.2018)

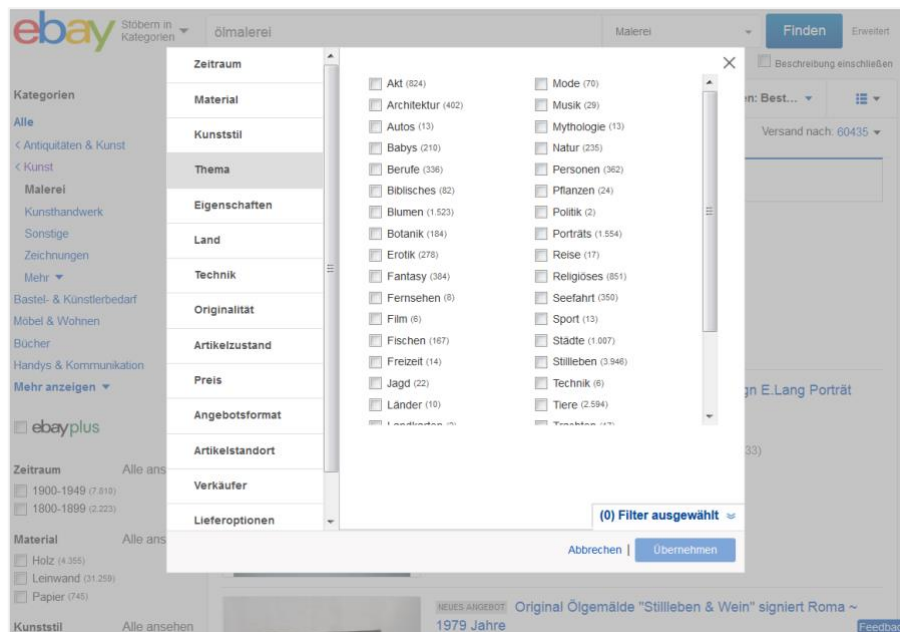
<sup>52</sup> siehe »All Categories« bei alibaba (URL: <https://www.alibaba.com/Products>), »Alle Kategorien« bei Amazon (URL: <https://www.amazon.de/gp/site-directory>) und bei ebay (URL: <https://www.ebay.de/v/allcategories>).

<sup>53</sup> Die Suchanfrage nach »Ölmalerei« mit der Voreinstellung »Alle Kategorien« bei ebay führt zu 67.000 Ergebnissen. Die Suchen wurden am 01.12.2018 durchgeführt. Die Ergebnismengen sind grob gerundet.



**Abbildung 2.9: Suchvorschläge mit Kategorien bei ebay<sup>54</sup>**

Auf der Ergebnisseite werden die Kategorien zum weiteren Einschränken oder erneuten Ausweiten der Suchtreffer angeboten. Dabei beschränkt sich die Sicht auf das Kategoriensystem nach jeder Auswahl einer Kategorie. Zum Verfeinern der Suchergebnisse gibt es verschiedene Filter, die ein Nutzer in Kombination mit den Kategorien aktivieren oder deaktivieren kann. Der Unterschied zum Navigieren über Kategorien besteht darin, dass beim Aktivieren bzw. Deaktivieren eines Filters die anderen Einstellungen zum Einschränken der Suchergebnisse erhalten bleiben<sup>55</sup>. Folgende Abbildung zeigt die Suchfilter in ebay, die zum Verfeinern von Ergebnissen zur Suche nach ›ölmalerie‹ in der Kategorie ›Malerei‹ angeboten werden. Der Themenfilter bietet hier 40 Schlagwörter an, die Nutzer auf der Suche nach Ölgemälden mit einem bestimmten Sujet oder Motiv weiterleiten können.



**Abbildung 2.10: Aktivierbare Filter bei ebay<sup>56</sup>**

<sup>54</sup> Quelle: ebay, URL: <http://www.ebay.de> (zitiert am: 01.12.2018)

<sup>55</sup> vgl. Hearst (2009, Kapitel 8.3).

<sup>56</sup> Quelle: ebay, URL: <http://www.ebay.de> (zitiert am: 01.12.2018)

## 2.2.4 Kulturerbeportale

Als Folge der strategischen Überlegungen der Europeana zur Verbesserung der Suche wurden im Portal Europeana Collections<sup>57</sup> seit 2016 elf »thematische Sammlungen« als navigierbare Suchunterstützung eingeführt, die den Sucheinstieg erleichtern sollen und auch als Filter auf der Ergebnisliste zum Verfeinern von Suchen angeboten werden<sup>58</sup>. Mit der Auswahl einer Sammlung können Nutzer den Gesamtbestand von aktuell (14.12.2018) 58 Millionen Kulturobjekten auf gesondert durchsuchbare Teilbestände beschränken. Auf der Startseite der Europeana bekäme der Nutzer aus dem Suchbeispiel dieser Arbeit eine erste Suchunterstützung durch die Zusammenstellung zum Thema Kunst. Das Navigationselement für die Sammlung »Art« wird in Form eines Menüpunktes (oben links) und eines verlinkten Vorschaubildes (unten rechts) angeboten:

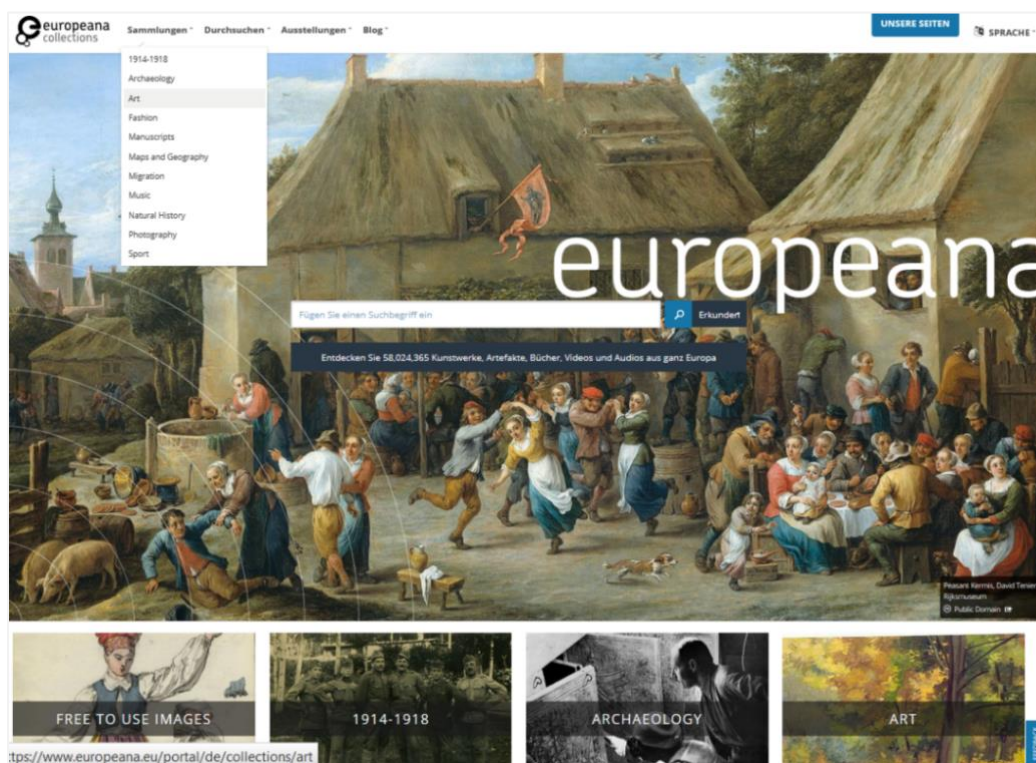


Abbildung 2.11: Sucheinstieg nach »Sammlungen« in Europeana Collections<sup>59</sup>

Welche Objekte in den Sammlungen durchsuchbar gemacht werden, wird durch die Metadaten der datengegebenden Einrichtungen bestimmt. Der »Europeana Publishing Guide Version 1.6« nennt zwei Kriterien, die für die Aufnahme eines Objektes in eine thematische Sammlung entscheidend sind: Der Bezug zum jeweiligen Thema muss über vorgegebene Schlagwörter kenntlich gemacht und die zugehörige Mediendatei über einen Link direkt zugänglich sein<sup>60</sup>. Der Vorteil dieses

<sup>57</sup> siehe <https://www.europeana.eu>

<sup>58</sup> vgl. Hill et al. (2016, S. 3).

<sup>59</sup> Quelle: Europeana Collections. URL: <https://www.europeana.eu/portal/de> (zitiert am 30.11.2018)

<sup>60</sup> siehe Scholz (2017, S. 12ff.).



Ansatzes ist, dass Nutzer den Gesamtbestand nach thematischen Gesichtspunkten segmentieren und innerhalb einer Auswahl an thematisch zusammengehörigen Objekten weitersuchen können. Dadurch sind zwar nicht alle, dafür aber relevante Objekte zum jeweiligen Thema leichter auffindbar. Durch die Auswahl der Sammlung »Art« wird der Suchraum beispielsweise auf 2,2 Millionen Objekte eingeschränkt:



**Abbildung 2.12: Einstiegseite der Sammlung »Europeana Art«<sup>61</sup>**

Ein Nachteil ist, dass die Sammlungen nicht das vollständige Themenspektrum der Europeana abbilden. Auch die 121 »Themenbereiche«, die unter den Menüpunkt »Durchsuchen« zugänglich sind, geben keinen Überblick über das Gesamtangebot und lassen keine Kriterien für ihre Auswahl erkennen<sup>62</sup>. Sie sind im Gegensatz zu den Kategorien der Verkaufskataloge auch nicht hierarchisch organisiert und bieten daher kaum Orientierung. Innerhalb der Sammlungsräume ermöglichen die scheinbar zufällig angebotenen »Themen« ebenfalls keine sinnvolle Filterung (in Abbildung 2.12 unten).

Auch die Objektdetailseiten von Europeana Collections bieten suchunterstützende Elemente an. In der Detailseite für das Objekt »Die Malerei« beispielsweise bekäme ein Nutzer in der Europeana Unterstützung bei seiner weiteren Suche nach der »Malerei als Thema«<sup>63</sup> (siehe Abbildung 2.13 unter »Klassifikationen«). Das im Anzeigeelement »Thema« verlinkte Schlagwort »symbolische Darstellungen, Allegorien und Embleme der Malerei; Pictura; Ripa: Pittura«<sup>64</sup> aus Iconclass löst eine

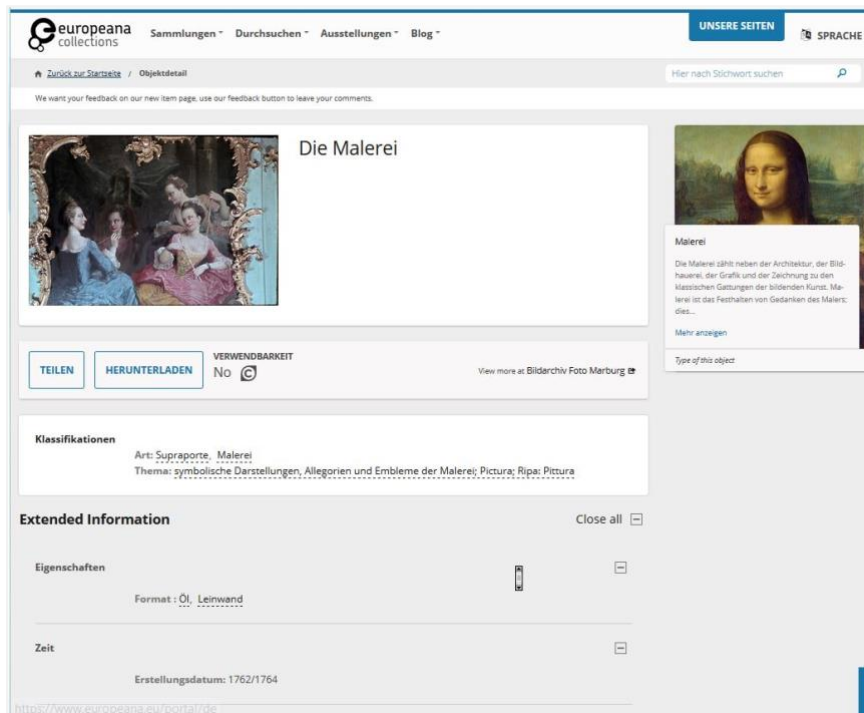
<sup>61</sup> Quelle: Europeana Collections. URL: <https://www.europeana.eu/portal/de> (zitiert am 01.12.2018)

<sup>62</sup> siehe <https://www.europeana.eu/portal/de/explore/topics.html>.

<sup>63</sup> vgl. Suchszenario 4 in Kapitel 2.1.3 für die Detailseite desselben Objektes in der DDB.

<sup>64</sup> Siehe: <https://www.europeana.eu/portal/de/search?q=what%3A%22symbolische+Darstellungen%2C+Allegorien+und+Embleme+der+Malerei%3B+Pictura%3B+Ripa%3A+Pittura%22>. Die Suche wurde am 01.12.2018 durchgeführt.

Suchanfrage aus, die aber lediglich 122 Skulpturen und Gemälde von zwei Datengebern zusammenführt, die genau dieses Themenschlagwort in den Metadaten indexiert haben<sup>65</sup>.



**Abbildung 2.13: Detailseite für das Objekt »Die Malerei« in Europeana Collections<sup>66</sup>**

Eine deutlich größere Trefferquote erhielt der Nutzer durch die Auswahl des Schlagwortes »Malerei« im Element »Art«, das mehr als 130.000 Objekte zusammenführt, darunter überwiegend Gemälde. Dieser verhältnismäßig gute Recall ist in der Verwendung eines kontrollierten Indexterms aus dem Europeana Vocabulary begründet<sup>67</sup>. In der Ergebnisseite wird dem Nutzer aber kein geeigneter Filter angeboten (links in Abbildung 2.14 unter »Suche verfeinern«), mit dem er seine Suche gezielt nach Objekten zum Thema »Malerei« einschränken könnte.

Dass Europeana ihren Nutzern derartige Filter zum Verfeinern von Suchergebnissen nach Themen und Objektarten anbieten möchte, verdeutlicht das RetrievalszENARIO »Browse by subjects and types«, das sich auf die Ergebnisse einer Personas-Studie stützt<sup>68</sup>. Als Voraussetzung für die Filter nennt das Dokument »Discovery – User scenarios and their metadata requirements«, dass entsprechende Metadatenelemente in den Datenlieferungen vorhanden und mit Werten aus spartenübergreifend gemeinsam genutzten Vokabularen belegt sein müssen<sup>69</sup>.

<sup>65</sup> In der Detailseite für dasselbe Objekt in der DDB ist das Schlagwort im Anzeigeelement »Bezug (Was)« nicht verlinkt.

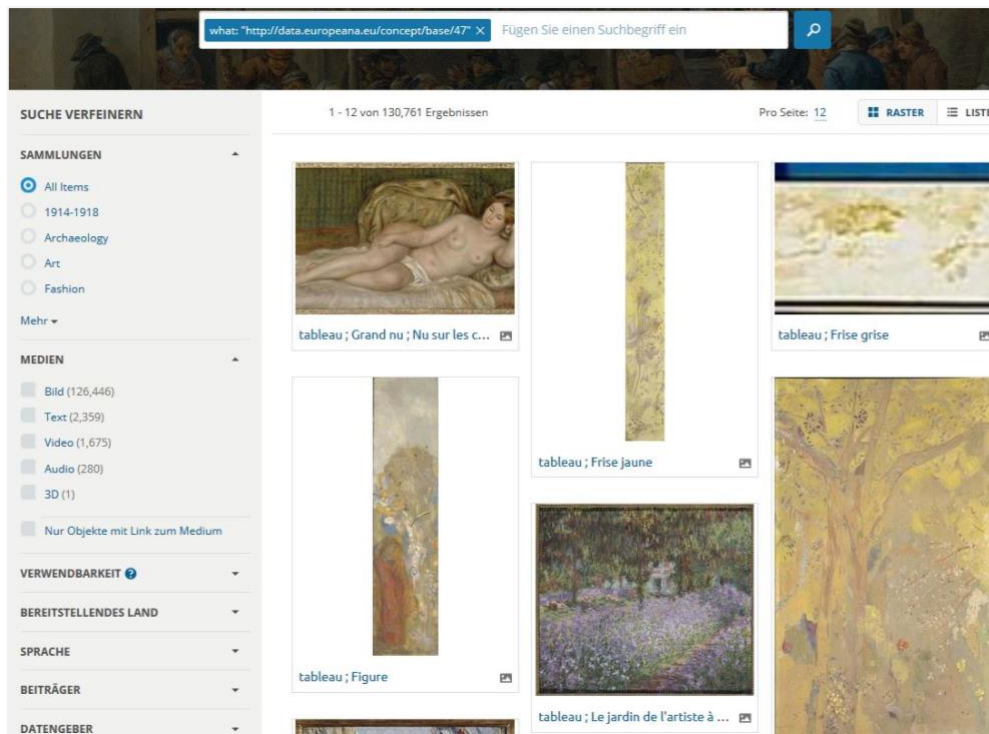
<sup>66</sup> Quelle: Europeana Collections, URL: [https://www.europeana.eu/portal/de/record/2064137/Museu\\_ProvidedCHO\\_Bildarchiv\\_Foto\\_Marburg\\_obj20784543.html](https://www.europeana.eu/portal/de/record/2064137/Museu_ProvidedCHO_Bildarchiv_Foto_Marburg_obj20784543.html) (letzte Aktualisierung: 15.06.2018)

<sup>67</sup> Europeana indexiert Objekte automatisch mit Begriffen aus Linked-Open-Data-Vokabularen, hier mit dem URI »<http://data.europeana.eu/concept/base/47>«, der für »Malerei« steht. Das Vorgehen der Datenanreicherung ist im »Europeana Semantic Enrichment Framework« von Manguinhas (17. November 2016) dokumentiert.

<sup>68</sup> »Scenario [...] As a user, I want to refine my results by concepts the CHO [Cultural Heritage Object] is about (as captured in the dc:subject element), and the concepts the CHO instantiates (i.e., dc:type)«, vgl. Charles, Isaac und Hill (2015, S. 23).

<sup>69</sup> vgl. Charles, Isaac und Hill (2015, S. 24 f.).





**Abbildung 2.14: Ergebnisseite für das Navigationselement ›Malerei‹ in Europeana Collections<sup>70</sup>**

## 2.2.5 Zusammenfassung und Schlussfolgerung

Informationsfilter werden seit den Anfängen des WWW zum Navigieren in thematisch und funktional unterschiedlichen Webanwendungen mit umfangreichen und heterogenen Datenbeständen angeboten. Zwar entspricht das Navigieren über komplexe und eindimensionale hierarchische Strukturen heutigen Nutzererwartungen nicht mehr, dennoch sind navigierbare Sucheinstiege über Kategoriensysteme wünschenswert. So hat sich das »Stöbern in Kategorien« in Portalen des elektronischen Handels erfolgreich etabliert. Der vorgestellte Ansatz der Europeana zeigt, dass auch im Kontext der Kulturerbeportale die Bedeutung von navigierbaren Sucheinstiegen über sachthematische Kategorien gesehen wird, unter anderem damit sich Nutzer einen strukturierten Überblick über das vielfältige Datenangebot verschaffen können. Ein navigierbarer Sucheinstieg mittels einer überschaubaren Systematik, deren Kategorien in verschiedenen Stellen des Suchprozesses mit facettenbasierten Filtern zum Verfeinern von Suchergebnissen kombiniert werden können, scheint heutzutage das Information-Retrieval in umfangreichen Datenbeständen optimal zu unterstützen. Darüber hinaus sind unterstützende Elemente, mit denen Nutzer ihre Suche gezielt ausweiten können, in Detailseiten von Objekten wünschenswert.

<sup>70</sup> Quelle: Europeana Collections, URL: <https://www.europeana.eu/portal/de/search?q=what%3A+%22http%3A%2F%2Fdata.europeana.eu%2Fconcept%2Fbase%2F47%22> (Zugriff am: 17.12.2018)

## 2.3 Mögliche Informationsfilter für die DDB

Ausgehend von den Erkenntnissen aus den vorangegangenen Analysen entwickelt dieses Kapitel Vorschläge für navigierbare Elemente, die in der Benutzeroberfläche des DDB-Portals zur Suchunterstützung angeboten werden könnten, um die Retrievalsituation zu verbessern. Wie in Kapitel 2.1.4 zusammengefasst, sollte der Stichwortfilter der DDB in drei unterschiedliche Facetten aufgeteilt werden, die in dieser Arbeit »Sachkategorie«, »Themenschlagwort« und »Objektgattung« genannt werden. Diese neu vorgeschlagenen Filter werden in den folgenden Abschnitten kurz definiert und erläutert. Danach werden jeweils die minimalen Voraussetzungen benannt, die an die DDB gelieferte Metadaten erfüllen müssen, damit sie als Elemente zum Navigieren nutzbar gemacht werden können.

### 2.3.1 Sachkategorie

#### **Definition:**

Ein Navigationselement, das Nutzer der DDB durch auswählbare Kategorien einer hierarchisch organisierten Systematik zu vordefinierten Ergebnismengen führt, die unter sachlich-systematischen Gesichtspunkten zusammengestellt wurden.

#### **Erläuterung:**

Mit dem Informationsfilter »Sachkategorie« sollen sich Nutzer einen ersten Überblick über das Angebot der DDB verschaffen können. Ähnlich wie bei den thematischen Sammlungen der Europeana sollten die Kategorien die Objekte des DDB-Bestandes in allgemeine thematische Gruppen einteilen und eine erste Vorauswahl zu Beginn der Suche ermöglichen. Die »Sachkategorien« der DDB sollten über wenige Hierarchiestufen ein Stöbern im Bestand und ein Blättern von allgemeinen zu speziellen Themengruppen unterstützen, zum Beispiel von einer Einstiegskategorie »Kunst und Kultur« zu einer Unterkategorie »Bildende Kunst« und von dort aus zur Unterkategorie »Malerei«. Die Kategorien könnten, wie in den Verkaufskatalogen, an verschiedenen Stellen im Suchprozess zum Navigieren angeboten werden, zum Beispiel (1) als Einstiegskategorien zu Beginn der Suche, (2) zum nachträglichen Verfeinern von Suchergebnismengen oder (3) zum gezielten Ausweiten von Suchen.

#### **Voraussetzungen:**

Um das Navigieren über Kategorien in der DDB zu ermöglichen, müssen die Objekte in ein hierarchisch strukturiertes Begriffssystem eingeordnet werden. An die DDB gelieferte Metadaten müssen daher mindestens folgende Voraussetzungen erfüllen:

1. In jedem Metadatenschema für die Datenlieferung an die DDB ist ein Element für die »Sachkategorie« definiert.
2. Das Element für die »Sachkategorie« ist in jedem Lieferdatensatz mindestens einmal vorhanden und enthält einen Wert für den Indexterm.

3. Der Indexterm entspricht der Kategorie eines hierarchisch organisierten Kategoriensystems, der Klasse eines Klassifikationssystems oder der Stelle einer Systematik.

### 2.3.2 Themenschlagwort

#### **Definition:**

Ein Navigationselement, mit dem Nutzer den Bestand der DDB anhand auswählbarer Schlagwörter, die das Thema der Kulturobjekte bezeichnen, filtern können.

#### **Erläuterung:**

Der Informationsfilter »Themenschlagwort« soll es Nutzern erlauben, ihre Suche auf das dargestellte, beschriebene oder abgebildete Thema zu konzentrieren. Mit der Auswahl des Schlagwortes »Malerei« beispielsweise sollten Nutzer u. a. zu gedruckten Werken mit dem Sachinhalt »Malerei«, zu Gemälden mit dem Sujet »Malerei« und zu Fotografien mit dem Motiv »Malerei« geführt werden. Die »Themenschlagwörter« können in den Ergebnislisten zum gezielten Einschränken von Suchergebnismengen und in den Detailseiten zum gezielten Ausweiten von Suchen angeboten werden.

#### **Voraussetzungen:**

Um ein Navigieren über »Themenschlagwörter« in der DDB zu ermöglichen, müssen die Objekte mit Schlagwörtern aus einem kontrollierten Vokabular erschlossen sein. An die DDB gelieferte Metadaten müssen daher mindestens folgende Voraussetzungen erfüllen:

1. In jedem Metadatenschema für die Datenlieferung an die DDB ist ein Element für das »Themenschlagwort« definiert.
2. Das Element ist in jedem Lieferdatensatz mindestens einmal vorhanden und enthält einen Wert für den Indexterm.
3. Der Indexterm entspricht einem Schlagwort aus einer Schlagwortliste oder Normdatei oder einem Deskriptor aus einem Thesaurus.

### 2.3.3 Objektgattung

#### **Definition:**

Ein Navigationselement, mit dem Nutzer den Bestand der DDB anhand auswählbarer Schlagwörter, die den generischen Typ der Kulturobjekte bezeichnen, filtern können.

#### **Erläuterung:**

Mit dem Informationsfilter »Objektgattung« sollen Nutzer ihre Suche auf bestimmte Arten von Objekten fokussieren können. Der Filter sollte zur Verfeinerung der Suche Schlagwörter anbieten, die je nach Art des beschriebenen Gegenstandes den Typ eines (1) schriftlichen Werkes (z. B.

›Lehrbuch‹), (2) visuellen Werkes (z. B. ›Gemälde‹), (3) auditiven Mediums (z. B. ›Schallplatte‹), (4) audiovisuellen Mediums (z. B. ›Dokumentarfilm‹) oder (5) dreidimensionalen Objektes (z. B. ›Skulptur‹) bezeichnen. Die »Objektgattungen« können als Navigationselemente in den Ergebnislisten zur Präzisierung von Suchen und in den Detailseiten zum gezielten Ausweiten von Suchen angeboten werden.

#### **Voraussetzungen:**

Um ein Navigieren über »Objektgattungen« in der DDB zu ermöglichen, müssen die Objekte mit Schlagwörtern aus einem kontrollierten Vokabular erschlossen sein. An die DDB gelieferte Metadaten müssen daher mindestens folgende Voraussetzungen erfüllen:

1. In jedem Metadatenschema für die Datenlieferung an die DDB ist ein Element für die »Objektgattung« definiert.
2. Das Element ist in jedem Lieferdatensatz mindestens einmal vorhanden und enthält einen Wert für den Indexterm.
3. Der Indexterm entspricht einem Schlagwort aus einer Schlagwortliste oder Normdatei oder einem Deskriptor aus einem Thesaurus.

## **2.4 Mögliche kontrollierte Vokabulare für die vorgeschlagenen Filter**

Dieses Kapitel fasst die wesentlichen Eigenschaften kontrollierter Vokabulare und ihre Bedeutung für das Information-Retrieval zusammen. Davon ausgehend werden Anforderungen formuliert, die diese im Kontext der DDB erfüllen sollten, um ein semantisches Retrieval über Informationsfilter zu unterstützen. Im Anschluss werden anhand dieser Kriterien einzelne Vokabulare vorgestellt, die als Navigationsinstrumente in der DDB in Frage kommen, und ihre jeweiligen Vor- und Nachteile werden benannt. Abschließend werden an einem konkreten Beispiel zwei verschiedene Szenarien vorgestellt, die illustrieren, wie durch die Nutzung von kontrolliertem Vokabular das Retrieval in der DDB verbessert werden könnte.

### **2.4.1 Kontrollierte Vokabulare und ihre Bedeutung für das Retrieval**

Kontrollierte Vokabulare, wie zum Beispiel Schlagwortlisten, Thesauri und Klassifikationen, werden in Bibliotheken, Archiven und Dokumentationseinrichtungen traditionell als Retrieval-instrumente eingesetzt. In digitalen Umgebungen können sie die Suche im Hintergrund unterstützen und an der Nutzeroberfläche als Werkzeuge zum Navigieren angeboten werden. Der Thesaurus-Standard ISO 25964-1:2011 definiert »controlled vocabulary« als »prescribed list of terms, headings or codes, each representing a concept« und ergänzt:

NOTE Controlled vocabularies are designed for applications in which it is useful to identify each concept with one consistent label, for example when classifying documents, indexing them and/or searching them. Thesauri, subject heading schemes and name authority lists are examples of controlled vocabularies. (ISO 25964-1:2011, 2.12)

Kontrollierte Vokabulare legen also einerseits verbindliche Terme für die Indexierung fest und dienen andererseits der terminologischen und begrifflichen Kontrolle. Terminologische Kontrolle meint, dass Begriffe und Bezeichnungen eindeutig aufeinander bezogen werden, indem (a) für jeden Begriff eine Bezeichnung (z. B. Vorzugsbenennung, Notation, Begriffsnummer) festgelegt wird, (b) Synonyme zusammengeführt und (c) Homonyme und Polyseme disambiguiert werden<sup>71</sup>. Unter begrifflicher Kontrolle wird die explizite Festlegung der Intension und Extension eines Begriffes durch die Definition und die Darstellung von Beziehungen zwischen den Begriffen verstanden. Diese paradigmatischen semantischen Relationen können in Retrievalprozessen unterschiedliche Funktionen erfüllen<sup>72</sup>:

- Hierarchierelationen unterstützen explorative Suchansätze und eine automatische Suchausweitung durch das Einbeziehen von Oberbegriffen und Unterbegriffen (Search Explode)<sup>73</sup>. Über die Begriffsleiter können Nutzer ihre Suche schrittweise einschränken oder ausweiten.
- Assoziationsrelationen unterstützen ebenfalls explorative Suchansätze, indem dem Nutzer thematisch verwandte Begriffe als weitere Suchterme vorgeschlagen werden.
- Äquivalenzrelationen können zur Suchausweitung eingesetzt werden, indem Synonyme automatisch als zusätzliche Zugriffspunkte in die Suche einbezogen oder dem Nutzer zum Auffinden eines passenden Suchterms angeboten werden.

Diese klassischen Thesaurusrelationen sind nicht in allen kontrollierten Vokabularen gleichermaßen vorhanden oder ausgearbeitet, weshalb für ein semantisches Retrieval über die hier vorgeschlagenen Informationsfilter nicht alle Typen von Vokabularen in Frage kommen.

Schlagwortlisten und Nomenklaturen sind in der Regel flache, nicht hierarchisch strukturierte Listen mit Termen für eine gebundene Indexierung. Im Gegensatz zu einer freien Verschlagwortung stützt sich die Indexierung also auf die endliche Liste verbindlicher Schlagwörter. Schlagwortlisten oder kleine Wertelisten für bestimmte Metadatenelemente haben oft keine Zugangsterminologie, während Nomenklaturen über gut ausgebaute Synonymrelationen verfügen. Beide Vokabulartypen sind als Navigationsinstrumente weniger gut geeignet, weil die notwendigen semantischen Relationen fehlen.

Als Navigationsinstrumente für die hier vorgeschlagenen Informationsfilter sind hierarchisch organisierte Vokabulare wie Klassifikationen und Thesauri besser geeignet, weil sie eine automatische Suchausweitung über Oberbegriffe und Unterbegriffe unterstützen.

Klassifikationen und Systematiken sind künstlichsprachig basierte Dokumentationssprachen, die zur Groberschließung vorgesehen sind und der Einordnung von Objekten in einen größeren Zusammenhang dienen. Die Indexterme sind hier Notationen, die durch Trunkieren eine einfache

---

<sup>71</sup> vgl. DIN 1463-1:1987-11-00 (S. 2).

<sup>72</sup> Eine tabellarische Übersicht zum Einsatz dieser Relationen in Retrievalprozessen findet sich im Aufsatz von Hubrich (2018, S. 282).

<sup>73</sup> vgl. ISO 25964-1:2011 (Clause 10.2, S. 58).

Ausweitung oder Einengung der Suche über die Hierarchie der Klassen ermöglichen. Die Notationen können auch als systematischer Zugang zu den Deskriptoren eines Vokabulars dienen. Als präkombinierte und zumeist monohierarchisch organisierte Instrumente sind klassifikatorische Systeme für eine Top-down-Navigation vom Allgemeinen zum Besonderen prinzipiell geeignet, sofern sie Standardanforderungen entsprechen<sup>74</sup>. Für den vorgeschlagenen Filter »Sachkategorie« kommen v. a. universelle Systeme in Frage, in denen das in ihnen repräsentierte Wissen nach Fachgebieten geordnet ist.

Thesauri sind Instrumente zur verbalen Feinerschließung mit Deskriptoren, die meist Einfachbegriffe oder Differenzbegriffe sind. Thesauri sind begriffsbasiert, dienen der punktuellen und präzisen Recherche und sind als post-kordinierende Instrumente für Schnittmengenbildung und facettierte Suchsysteme angelegt. Die Deskriptoren eignen sich als Filterwerte für Facetten, sofern der Thesaurus Standardanforderungen entspricht<sup>75</sup>: eindeutige Identifizierung eines Begriffes, terminologische Kontrolle und logische Hierarchiestrukturen. Für die vorgeschlagenen Filter »Themenschlagwort« und »Objektgattung« kommen Thesauri mit breiter inhaltlicher Abdeckung besonders in Frage.

In Informationsräumen wie der DDB, die durch die Heterogenität der in ihr zusammengeführten objektbeschreibenden Metadaten geprägt sind, können einzelne kontrollierte Vokabulare allerdings nur bedingt Suchunterstützung leisten, weil sie jeweils nur auf einen Teilbestand zugreifen. Ein Ansatz, diese Heterogenität zu überbrücken, ist das Erstellen sogenannter Mappings, qualifizierter Verbindungen zwischen den bei der Indexierung verwendeten und in den Vokabularen festgeschriebenen Begriffen. Durch diese Mapping-Relationen werden unterschiedlich erschlossene Bestände gemeinsam zugänglich, weil für Abfragen die Begriffe des jeweils anderen Systems einbezogen werden können<sup>76</sup>. Gerade integrierende Rechercheportale wie die DDB können von diesem Ansatz profitieren.

Einen wesentlichen Beitrag für ein integrierendes semantisches Retrieval leisten in diesem Zusammenhang die Technologien des Semantic Web und die Prinzipien von »Linked Open Data«. Dazu werden Vokabulare mit ihren semantischen Relationen in maschinenlesbaren Datenformaten auf der Basis von RDF<sup>77</sup> modelliert und über Webservices kostenfrei anderen Webanwendungen zur Verfügung gestellt. Die Begriffe werden durch persistente, im WWW eindeutig referenzierbare Identifikatoren, Uniform Resource Identifiers (URIs), repräsentiert. Durch die Verwendung von standardisierten Datenmodellen, wie zum Beispiel dem Simple Knowledge Organization System (SKOS), werden den Relationen explizit maschinenlesbare Eigenschaften zugeordnet, durch die automatisch Inferenzen gezogen werden können. Auf diese Weise können zum Beispiel Relationen für eine Suchausweitung und das Navigieren über Hierarchien in der DDB

---

<sup>74</sup> DIN 32705:1987 zur Erstellung und Weiterentwicklung von Klassifikationssystemen.

<sup>75</sup> ISO 25964-1:2011, zumindest DIN-Norm DIN 1463-1:1987-11-00 zur Erstellung und Weiterentwicklung von Thesauri.

<sup>76</sup> vgl. Hubrich (2018, S. 267 f.).

<sup>77</sup> Resource Description Framework

nutzbar gemacht werden, sofern die Kulturobjekte mit geeigneten kontrollierten Vokabularen erschlossen sind, die als Linked Open Data (LOD) veröffentlicht sind.

## **2.4.2 Anforderungen an kontrollierte Vokabulare im Kontext der DDB**

### **(1) Kompatibilität zu Standards der Dokumentation und Wissensorganisation**

Vokabulare, die als Navigationsinstrumente in der DDB zum Einsatz kommen sollen, müssen zumindest die in Vokabularstandards empfohlenen Voraussetzungen an eindeutig identifizierbare Deskriptoren erfüllen. Wünschenswert sind außerdem (a) Definitionen für die Begriffe, (b) ein standardisiertes Datenmodell, das die Struktur des Vokabulars abbildet und (c) begleitende Anwendungsregeln für die Indexierung.

### **(2) Entwicklung und Pflege durch anerkannte Organisationen**

Die Vokabulare sollten von im BAM-Bereich anerkannten Organisationen herausgegeben und möglichst kooperativ von verschiedenen Institutionen weiterentwickelt und gepflegt werden. Durch die Organisationsform sollten Datenerfassung, Datenhaltung und Datenpflege langfristig sichergestellt und Möglichkeiten der schnellen Aktualisierung gegeben sein.

### **(3) Breite inhaltliche Abdeckung**

Die Vokabulare sollten universell und thematisch umfassend sein. Geeignet sind Vokabulare, die Wissensgebiete und/oder Objektgattungen vor allem in der Breite abdecken.

### **(4) Breite Nutzung**

Die Vokabulare sollten weit verbreitet und von möglichst vielen kulturellen Einrichtungen zur Erschließung ihrer Objekte genutzt werden. Besonders geeignet sind Vokabulare, die über Spartengrenzen hinweg verwendet und von Projekten zur Zusammenführung des digitalen Kulturerbes empfohlen werden. Der Einsatz des Vokabulars als Retrievalinstrument in Webanwendungen mit heterogenen Datenbeständen ist ein zusätzlicher Indikator für dessen Eignung.

### **(5) Semantische Interoperabilität**

Besonders geeignet sind Begriffssysteme, die über standardkonforme Mappings<sup>78</sup> bzw. Konkordanzen mit anderen Vokabularen, die von BAM-Institutionen zur Indexierung verwendet werden, verbunden sind.

### **(6) Freier Zugang und Linked Open Data**

Die Vokabulare sollten im Web kostenfrei zugänglich und unter den Bedingungen einer freien Lizenz nutzbar sein. Darüber hinaus sollten sie in einem standardisierten RDF-basierten Datenformat publiziert und die Begriffe durch URIs repräsentiert sein.

---

<sup>78</sup> Ein Standard zur Herstellung semantischer Interoperabilität zwischen Thesauri und anderen Vokabularen ist ISO 25964-2:2013.

### 2.4.3 Dewey-Dezimalklassifikation (DDC) und DDC-Sachgruppen

Die Dewey-Dezimalklassifikation (DDC) ist eine Universalklassifikation, die von dem amerikanischen Bibliothekar Melvil Dewey konzipiert und 1876 in den USA erstmals veröffentlicht wurde. Sie ist weltweit das am meisten genutzte Klassifikationssystem und findet seit ihrer Übersetzung ins Deutsche<sup>79</sup> im Jahr 2005 auch im deutschsprachigen Raum mehr Anwendung.

**(1)** Die DDC ist überwiegend monohierarchisch organisiert und nach dem Dezimalprinzip aufgebaut, das Erweiterungsmöglichkeiten der Begriffsleitern bietet, aber keine Erweiterungen der Begriffsreihen über zehn Klassen hinaus gestattet. Diverse Fachgebiete sind daher häufig an einer unerwarteten Stelle in der Hierarchiestruktur zu finden. Vor allem aber weist die DDC Defizite in der hierarchischen Organisation auf, weil die Merkmale der Unterteilung, die sogenannten Klasseme, nicht einheitlich gewählt sind. Eine logische Abfolge untergeordneter Klassen ist oft nicht eingehalten und auf ein und derselben Hierarchieebene können Fachgebiete mit unterschiedlichem Abstraktionsniveau, Spezialisierungsgrad und Begriffsumfang nebeneinander stehen<sup>80</sup>.

**(2)** Die DDC wird von der Library of Congress und dem OCLC Online Computer Library Center in Abstimmung mit den internationalen Anwendergruppen aktualisiert, die Lizenz liegt bei OCLC. Die Deutsche Nationalbibliothek (DNB) hat die Pflege und Weiterentwicklung der deutschsprachigen DDC übernommen.

**(3)** Die DDC wurde primär aus US-amerikanischer Perspektive entwickelt und die oberen Gliederungsstufen spiegeln eine überholte Systematik der Wissenschaften wider<sup>81</sup>. Durch stetige Überarbeitung versucht die DDC, moderne Themen und Wissenschaftsdisziplinen an einer fachlich zutreffenden Stelle einzufügen.

**(4)** Die Nutzung der DDC verzeichnete in Deutschland einen wesentlichen Anstieg nach ihrer Einführung durch die Nationalbibliothek<sup>82</sup>. Seit 2004 sind die Reihen der Deutschen Nationalbibliografie und des Neuerscheinungsdienstes nach Sachgruppen gegliedert, die überwiegend den oberen beiden Hierarchieebenen der DDC entsprechen und circa 100 Klassen umfassen<sup>83</sup>. Die DDC wird zudem in Webanwendungen, z. B. in virtuellen Fachbibliotheken, verwendet<sup>84</sup>.

---

<sup>79</sup> Die Übersetzung wurde im Rahmen des DFG-Projektes »DDC-Deutsch« (2002-2005) von der DNB in Zusammenarbeit mit der Fachhochschule Köln realisiert. Für die DDC-Deutsch hat die DNB eine Informationswebsite eingerichtet, siehe <http://www.ddc-deutsch.de>.

<sup>80</sup> Zu allen genannten Kritikpunkten an der DDC vgl. Umlauf (1999, Kapitel 3.1).

<sup>81</sup> vgl. Alex (16. Mai 2018).

<sup>82</sup> Zu diesem Ergebnis kam das DFG-Projekt »Coli-Conc« (2015-2017) im Rahmen einer Online-Umfrage zu Sacherschließungsinstrumenten und Konkordanzprojekten in Bibliotheken und anderen informationsvermittelnden Einrichtungen im deutschsprachigen Raum. Demnach wird die DDC von 26 Prozent der Einrichtungen benutzt und liegt damit zehn Prozentpunkte hinter der am häufigsten verwendeten Regensburger Verbundklassifikation, vgl. Balakrishnan und Agne (2016, S. 10).

<sup>83</sup> siehe Deutsche Nationalbibliothek (1. Januar 2011). Für die Vergabe der DDC-Sachgruppen gibt die DNB einen Leitfaden heraus, siehe Deutsche Nationalbibliothek (1. Juni 2014).

<sup>84</sup> Eine Liste der Bibliotheken und Projekte in Deutschland, die mit der DDC arbeiten, findet sich unter: [http://www.dnb.de/Subsites/ddcdeutsch/DE/Anwendung/BibliothekenProjekte/bibliothekenProjekte\\_node.html](http://www.dnb.de/Subsites/ddcdeutsch/DE/Anwendung/BibliothekenProjekte/bibliothekenProjekte_node.html). Auf die Nutzung der DDC im EU-Projekt Renardus wurde in Kapitel 2.2.2 dieser Arbeit kurz eingegangen.



(5) Die DDC-Sachgruppen der deutschen Nationalbibliografien sowie einzelne Sektionen der DDC sind über Konkordanzen mit der Regensburger Verbundklassifikation und der Basisklassifikation verbunden, die in verschiedenen Datenformaten über die Konkordanz-Datenbank des Gemeinsamen Bibliotheksverbundes (GBV) bezogen werden können<sup>85</sup>.

(6) Die DDC ist nur eingeschränkt kostenfrei im Web zugänglich und mittlerweile nicht mehr als LOD-Vokabular verfügbar<sup>86</sup>. Der im Browsingtool WebDewey Search<sup>87</sup> angebotene Teil der »DDC Deutsch« kann unter den Bedingungen der Creative Commons-Lizenz »Namensnennung - Nicht-kommerziell - Keine Bearbeitung 3.0 Deutschland« (CC BY-NC-ND 3.0 DE)<sup>88</sup> unter Nennung des Rechteinhabers OCLC genutzt werden.

#### Nutzungsoption für die DDB:

Aufgrund der tiefen Schachtelung und der schlechten Vorhersehbarkeit des Begriffsumfangs der untergeordneten Klassen ist die DDC für eine Top-down-Navigation weniger gut geeignet. Dennoch könnte sie wegen ihres universellen Charakters und ihrer breiten Anwendung genutzt werden, um die heterogen erschlossenen Bestände der DDB auf einer weniger feingranularen Ebene zusammenzuführen. Auf dem Niveau der DDC-Sachgruppen könnte zumindest ein grober Sucheinstieg über Sachkategorien realisiert werden, vorausgesetzt, das Vokabular ist als LOD-Vokabular verfügbar.

Das folgende Beispiel zeigt einen möglichen Sucheinstieg in der DDB über die Klassenhierarchie in der »DDC-Deutsch« bis zur Notation ›751.45‹ (Thema: Ölmalerei). Bei der Nutzung der DDC-Sachgruppen würde die Begriffsleiter bei der Notation ›750‹ (Thema: Malerei) enden. Die untergeordneten DDC-Klassen könnten jedoch automatisch in die Suchergebnismenge für ›750‹ einbezogen werden, indem die Notationen an der dritten Systemstelle trunkiert würden.

Haupttafeln	
Notation	Thema
	<a href="#">Haupttafeln</a>
<a href="#">700</a>	<a href="#">Künste &amp; Unterhaltung</a>
<a href="#">750</a>	<a href="#">Malerei</a>
<a href="#">751</a>	<a href="#">Techniken, Verfahren, Geräte, Ausstattung, Materialien, Formen</a>
<a href="#">751.4</a>	<a href="#">Techniken und Verfahren</a>
<b>751.45</b>	<b>Ölmalerei</b>

Abbildung 2.15: Begriffsleiter für die Klasse ›751.45‹ in der »DDC Deutsch«<sup>89</sup>

<sup>85</sup> siehe <http://coli-conc.gbv.de/concordances/>.

<sup>86</sup> Der Webservice Dewey.info wurde von OCLC eingestellt, vgl. Online Computer Library Center (9. Juni 2015).

<sup>87</sup> siehe <https://deweysearchde.pansoft.de/webdeweysearch/>.

<sup>88</sup> siehe <https://creativecommons.org/licenses/by-nc-nd/3.0/de>.

<sup>89</sup> Quelle: WebDeweySearch, URL: <https://deweysearchde.pansoft.de/webdeweysearch/> (zitiert am 18.11.2018).

#### 2.4.4 Gemeinsame Normdatei (GND)

Die GND ist eine hybride Normdatei für Individualnamen und Sachbegriffe. Sie ging im Jahr 2012 aus einer Integration von vier bibliothekarischen Normdateien hervor, darunter die Schlagwortnormdatei (SWD), die im heutigen GND-Datenmodell dem Entitätentyp der Sachbegriffe entspricht.

**(1)** Der GND-Bereich der Sachbegriffe weist wesentliche Merkmale eines Thesaurus auf mit Einstiegsvokabular, Ober- und Unterbegriffen sowie assoziativ verbundenen Deskriptoren. Weil die SWD ursprünglich als Nomenklatur angelegt war, sind in der GND die hierarchischen Relationen noch nicht vollständig ausgearbeitet und Disambiguierungen nicht konsequent für alle Homonyme durchgeführt. Nach den kürzlich aktualisierten »Regeln für die Schlagwortkatalogisierung« (RSWK) wird eine durchgängige hierarchische Relationierung der Sachschlagwörter für die Navigation in der GND angestrebt<sup>90</sup>. Die Vergabe von identifizierenden Zusätzen ist auch weiterhin nicht für alle Homonyme vorgesehen<sup>91</sup>.

**(2)** Die GND wird von der DNB, den bibliothekarischen Verbünden und den ihnen angeschlossenen Bibliotheken, der Zeitschriftendatenbank und weiteren Einrichtungen geführt. Die Pflege erfolgt kooperativ von allen Teilnehmern mit unterschiedlichen Berechtigungen. Für die Abstimmung der Anwendung gemeinsamer Regeln und Standards sind der Standardisierungsausschuss und die in dessen Auftrag handelnde Arbeitsstelle für Standardisierung zuständig. Seit 2017 öffnet sich die GND für die Mitarbeit von Kultursparten außerhalb des Bibliothekswesens<sup>92</sup>. Die dafür notwendige Erneuerung der Organisationsform wird aktuell im Rahmen des DFG-Projektes »GND für Kulturdaten (GND4C)« (2018-2020) erarbeitet<sup>93</sup>.

**(3)** Die GND ist ein universelles Vokabular mit breiter inhaltlicher Abdeckung. Im Dezember 2017 verzeichnete sie rund 211.000 Sachbegriffe<sup>94</sup> aus allen einschlägigen Fachgebieten.

**(4)** Die GND wird traditionellerweise von Bibliotheken für die verbale inhaltliche Erschließung ihrer Bestände verwendet, zunehmend aber auch von Archiven, Museen, Kooperationsprojekten und in zahlreichen Webanwendungen genutzt<sup>95</sup>.

**(5)** Die GND ist mit anderen Vokabularen verbunden. Es gibt z. B. Konkordanzen zur amerikanischen Normdatei Library of Congress Subject Headings (LCSH) und zur französischen Normdatei RAMEAU. Für die Unterstützung explorativer Suchansätze in heterogenen Informationsräumen sind v. a. die

---

<sup>90</sup> vgl. Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (März 2017, § 12, S. 60).

<sup>91</sup> Nach der RSWK-Regel § 10,2 und der Erfassungshilfe EH-S-01 entfällt der identifizierende Zusatz i. d. R. bei der bevorzugten Bezeichnung, weil sie die Grundbedeutung eines Sachbegriffes repräsentiert, vgl. Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (März 2017, § 10,2, S. 56); Deutsche Nationalbibliothek (1. März 2017).

<sup>92</sup> vgl. Deutsche Nationalbibliothek (2017, S. 14).

<sup>93</sup> vgl. Niggemann et al. (2017, S. 3f.).

<sup>94</sup> Die Zahl stammt aus dem DNB-Jahresbericht 2017, vgl. Deutsche Nationalbibliothek (2018, S. 51).

<sup>95</sup> Referenzen zu einschlägigen Projekten und Webanwendungen finden sich im Antrag für das GND4C-Projekt, Niggemann et al. (2017, S. 4-8).

Verbindungen zwischen den Sachschlagwörtern der GND und Notationen der DDC<sup>96</sup> sowie den Sachgruppen der GND-Systematik<sup>97</sup> relevant.

(6) Verschiedene Webzugänge ermöglichen eine kostenfreie Recherche in der GND<sup>98</sup>. Alle GND-Datensätze stehen als RDF über den Linked-Data-Service der DNB<sup>99</sup> kostenfrei unter der Creative-Commons-Lizenz »CC0 1.0 Universal (CC0 1.0) Public Domain Dedication«<sup>100</sup> zur Verfügung. Die Verknüpfungen zu den DDC-Notationen und GND-Sachgruppen sind darin als integrierter Teil der GND-Datenbasis enthalten. Den RDF-Repräsentationen der GND-Datensätze liegt die GND-Ontologie (GNDO)<sup>101</sup> zugrunde.

#### **Nutzungsoption für die DDB:**

Die GND ist als LOD-Vokabular, aufgrund ihrer breiten inhaltlichen Abdeckung und spartenübergreifenden Perspektive sowie mit ihren semantischen Verknüpfungen besonders geeignet, als gemeinsames Bindeglied für das Indexat aggregierter Bestände in der DDB genutzt zu werden. Die Deskriptoren könnten als Filterwerte der vorgeschlagenen Facetten »Themenschlagwort« und »Objektgattung« ein post-kordinierendes Retrieval unterstützen, die verbundenen DDC-Notationen und GND-Sachgruppen wiederum könnten als Grundlage für eine Top-down-Navigation über »Sachkategorien« dienen. Für derartige Nutzungsszenarien müsste die GND jedoch als spartenübergreifendes und thesaurusbasiertes Erschließungsinstrument ausgebaut und von möglichst vielen Einrichtungen zur Indexierung ihrer Bestände verwendet werden.

Das folgende Beispiel zeigt den Sachbegriff »Ölmalerei« mit seinen Relationen im Katalog der Deutschen Nationalbibliothek<sup>102</sup>. Über die Relationen zur DDC-Notation »751.45« (Ölmalerei) und GND-Sachgruppe »13.3« (Malerei) könnten mit diesem Deskriptor indexierte Kulturobjekte in »Sachkategorien« der DDB eingeordnet werden. An diesem Beispiel wird auch deutlich, dass die GND ein präzises Retrieval in der DDB derzeit nicht optimal unterstützt. »Ölgemälde« ist hier als Synonym erfasst, das in der Museumsdokumentation<sup>103</sup> für die Indexierung des Objekttyps verwendet wird<sup>104</sup>. In der Suchergebnismenge für »Ölmalerei« in der DDB würden folglich nicht nur Druckwerke über die Maltechnik, sondern auch Ölgemälde der Museen enthalten sein. Ohne eine Disambiguierung dieses GND-Sachbegriffes könnte »Ölgemälde« von den Museen nicht als

---

<sup>96</sup> Diese Verbindungen wurden in der SWD durch intellektuelle Mappings im Rahmen des CrissCross-Projektes (2006-2010) angelegt und in der GND weitergeführt. Das Projekt ist im Aufsatz von Hubrich (2018, S. 285 f.) kurz beschrieben.

<sup>97</sup> Die dreistufige GND-Systematik verfügt über 484 Sachgruppen und 37 Klassen auf der obersten Ebene. Sie dient vorrangig als systematisches Zugangsvokabular zu den GND-Deskriptoren. Für die Notationenvergabe gibt die DNB einen Leitfaden heraus, siehe Deutsche Nationalbibliothek (11. November 2011).

<sup>98</sup> z. B. Katalog der Deutschen Nationalbibliothek (URL: <https://portal.dnb.de/opac.htm>), OGND (URL: <http://ognd.bsz-bw.de>), WebGND (URL: <http://gnd.eurospider.com>) und lobid-gnd (URL: <https://lobid.org/gnd>).

<sup>99</sup> siehe [http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata\\_node.html](http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata_node.html).

<sup>100</sup> siehe <https://creativecommons.org/publicdomain/zero/1.0/>.

<sup>101</sup> siehe <http://d-nb.info/standards/elementset/gnd>.

<sup>102</sup> Quelle: Katalog der Deutschen Nationalbibliothek, URI <http://d-nb.info/gnd/4043238-5>, Zugriff am 18.11.2018.

<sup>103</sup> Die Anforderungen an Sachbegriffe in der GND zur Beschreibung von Museumsobjekten wird aktuell im Rahmen einer Fallbeispielanalyse im GND4C-Projekt erhoben, vgl. Niggemann et al. (2017, S. 13f.).

<sup>104</sup> Zur Quasisynonymie in der GND siehe zum Beispiel Hubrich (2018, S. 271).

Indexterm für den Objekttyp verwendet und demzufolge nicht als Filterwert in der vorgeschlagenen DDB-Facette »Objektgattung« angeboten werden.

<b>Link zu diesem Datensatz</b>	<a href="http://d-nb.info/gnd/4043238-5">http://d-nb.info/gnd/4043238-5</a>
<b>Sachbegriff</b>	Ölmalerei
<b>Quelle</b>	M
<b>Erläuterungen</b>	Definition: Wird i.d.R. nur verwendet, wenn die Maltechnik behandelt ist.
<b>Synonyme</b>	Ölgemälde
<b>Oberbegriffe</b>	Malerei
<b>DDC-Notation</b>	751.45
<b>Systematik</b>	13.3 Malerei
<b>Typ</b>	Allgemeinbegriff (saz)
<b>Andere Normdaten</b>	RAMEAU: Peinture à l'huile

**Abbildung 2.16: Sachbegriff »Ölmalerei« im Katalog der Deutschen Nationalbibliothek**

Als Desiderat für eine spartenübergreifende Erschließung mit der GND müssen auch fehlende Typisierungen der Sachbegriffe genannt werden. Während zum Beispiel »Lehrbuch« als normierter Begriff für die Beschreibung der »Formangabe zur Art des Inhalts (saf)« einer Ressource angelegt ist<sup>105</sup>, fehlt bei »Gemälde« der entsprechende Entitätentyp, der aussagt, dass dieser Sachbegriff zur Indexierung der Art des Objektes dient, was v. a. für die Museumsdokumentation relevant wäre<sup>106</sup>.

## 2.4.5 Art & Architecture Thesaurus (AAT)

Der AAT ist ein polyhierarchisch strukturierter, facettierter und multilingualer Thesaurus für die Erschließung und das Retrieval von Objekten der Kunst, Architektur und materiellen Kultur. Eine deutsche Fassung wurde im Jahr 2012 vom Institut für Museumsforschung begonnen<sup>107</sup>.

(1) Der AAT ist weitgehend konform mit dem Thesaurusstandard ISO 25964-1. Die Deskriptoren sind in acht Hauptfacetten<sup>108</sup> organisiert. Die Vorteile des AAT liegen v. a. in der Eindeutigkeit und Präzision seiner Deskriptoren, den durchgängig verfügbaren Definitionen, den differenzierten Assoziationsrelationen und der logischen hierarchischen Struktur der Begriffe. Einen zusätzlichen Vorteil bieten die begleitenden Regelwerke »Categories for the Description of Works of Art« (CDWA)<sup>109</sup> und »Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images« (CCO)<sup>110</sup>.

(2) Der vom Getty Research Institute herausgegebene Thesaurus wird seit Ende der 1970er Jahre gemeinschaftlich von 200 Beiträgern fortlaufend weiterentwickelt<sup>111</sup>.

<sup>105</sup> Quelle: Katalog der Deutschen Nationalbibliothek, URI <http://d-nb.info/gnd/4123623-3>, Zugriff am 26.01.2019.

<sup>106</sup> Quelle: Katalog der Deutschen Nationalbibliothek, URI <http://d-nb.info/gnd/4122164-3>, Zugriff am 26.01.2019.

<sup>107</sup> Die Übersetzungsarbeiten erfolgten im Rahmen des vom Bundesbeauftragten für Kultur und Medien (BKM) geförderten Projektes »AAT Deutsch« (2012-2014), siehe <http://www.aat-deutsch.de/>.

<sup>108</sup> Assoziierte Begriffe, Begriffe physischer Merkmale, Stile und Epochen, Akteure, Aktivitäten, Material, Objekte und Markennamen

<sup>109</sup> siehe The J. Paul Getty Trust (23. Oktober 2017).

<sup>110</sup> siehe Visual Resources Association (2006).

<sup>111</sup> Eine Liste der Beiträger findet sich unter: <http://www.getty.edu/research/tools/vocabularies/contributors.html>.

(3) Der AAT wird vor allem in der Museumsdokumentation<sup>112</sup> und von kunsthistorischen Archiven für die Objekterschließung verwendet, in zahlreichen Webanwendungen des digitalen Kulturerbes eingesetzt und von entsprechenden Projekten zur Anwendung empfohlen<sup>113</sup>. Auch das Regelwerk »Describing Archives: A Content Standard« (DACS) empfiehlt den AAT als primäre Quelle für die normierte Erschließung von Archivalientypen<sup>114</sup>.

(4) Der AAT deckt die Bereiche der Kunst, Architektur und materiellen Kultur breit gefächert ab. Im Juli 2017 verzeichnete der AAT circa 60.000 Datensätze, 375.000 Bezeichnungen und 15.420 Kandidatenterme<sup>115</sup>.

(5) Der AAT hat keine nennenswerten Konkordanzen zu anderen Vokabularen.

(6) Der AAT ist über ein Recherchetool im Web frei zugänglich<sup>116</sup> und seit 2014 als LOD-Vokabular unter der »Open Data Commons Attribution License« (ODC-By) 1.0<sup>117</sup> veröffentlicht. Die RDF-Repräsentation der Deskriptoren ist kompatibel mit SKOS und seinen Erweiterungen.

#### **Nutzungsoption für die DDB:**

Der AAT wäre für ein facettenbasiertes Retrieval in der DDB vor allem wegen seiner kategoriell klar unterschiedenen Begriffe und der logischen Hierarchiestruktur gut geeignet. So ist beispielsweise ›oil paintings‹ im AAT zum einen als visuelles Werk definiert<sup>118</sup> und der Objektfacette zugeordnet und zum anderen als Maltechnik definiert<sup>119</sup> und der Aktivitätenfacette zugeordnet. Im Deutschen würde der AAT-Deskriptor ›oil paintings‹ mit dem identifizierenden Zusatz »(visual works)« dem Schlagwort ›Ölgemälde‹ und der AAT-Deskriptor ›oil paintings‹ mit dem Zusatz »(technique)« dem Schlagwort ›Ölmalerei‹ entsprechen. Weil der AAT Polyseme und Homonyme durchgängig disambiguiert, wie im beschriebenen Beispiel dargelegt, könnte er ein präzises Retrieval in der DDB gut unterstützen, vorausgesetzt, die entsprechenden Deskriptoren wären ins Deutsche übersetzt.

Die Begriffsleiter der »Objects Facet« des AAT könnte in der vorgeschlagenen DDB-Facette »Objektgattung« zum gezielten Einschränken und Ausweiten von Suchergebnissen zum Einsatz kommen. Zusätzlich könnten diejenigen »Guide Terms« des AAT, die das Merkmal der hierarchischen Untergliederung explizit benennen, die »Node Labels«<sup>120</sup>, als Unterfacette innerhalb der DDB-Facette genutzt werden. Im folgendem Beispiel, das die Begriffsleiter für den AAT-

---

<sup>112</sup> Zur Verwendung des AAT in der Museumsdokumentation siehe die »Ergebnisse der Online-Umfrage zur Nutzung von kontrolliertem Vokabular in Museen im Herbst 2016« von Haffner und Schweibenz (2017).

<sup>113</sup> z. B. von Europeana, siehe Charles und Devarenne (4. Dezember 2014).

<sup>114</sup> vgl. Society of American Archivists (2013, S. XXIV, Documentary Forms).

<sup>115</sup> vgl. The J. Paul Getty Trust (20. August 2018).

<sup>116</sup> siehe <http://www.getty.edu/research/tools/vocabularies/aat/index.html>.

<sup>117</sup> siehe <https://opendatacommons.org/licenses/by/1-0/>.

<sup>118</sup> siehe <http://vocab.getty.edu/page/aat/300033799>

<sup>119</sup> siehe <http://vocab.getty.edu/page/aat/300178684>

<sup>120</sup> Dies ist die nach dem ISO-Standard 25964-1:2011 gebräuchliche Bezeichnung.

Deskriptor ›oil paintings (visual works)‹ darstellt, ist solch ein »Node Label« beispielsweise »<paintings by material or technique>«.



**Abbildung 2.17: Begriffsleiter für den AAT-Deskriptor ›oil paintings (visual works)‹**<sup>121</sup>

## 2.4.6 Zusammenfassung und Schlussfolgerung

Kontrollierte Vokabulare spielen für gute Retrievalergebnisse im Sinne von Vollständigkeit und Genauigkeit eine entscheidende Rolle. Für eine Suchunterstützung mit Informationsfiltern in der DDB sind eindeutig identifizierbare Deskriptoren aus verbreiteten kontrollierten Vokabularen mit gut ausgearbeiteten semantischen Relationen eine notwendige Voraussetzung, damit Zusammenhänge zwischen den Kulturobjekten aufgedeckt und für eine Suchausweitung und das Navigieren über Hierarchien nutzbar gemacht werden können.

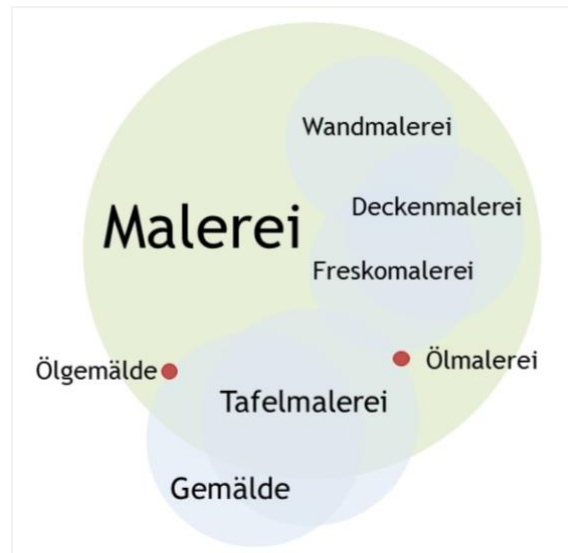
Die folgenden Venn-Diagramme zeigen, wie durch die gemeinsame Nutzung von kontrolliertem Vokabular die Suchergebnismengen in der DDB bezüglich der Vollständigkeit auf effektive Weise verbessert werden könnten. Im Beispiel wurde die GND als mögliches spartenübergreifendes semantisches Bindeglied für den Kulturerbe-Bereich verwendet. Es nimmt Bezug auf die im Venn-Diagramm der Abbildung 2.7 dargestellte Ausgangslage zum Informationsverlust bei der Suche in der DDB. Ausgewertet wurden die semantischen Relationen in den RDF-Repräsentationen derjenigen GND-Deskriptoren, die den Stichwörtern der DDB am besten entsprechen<sup>122</sup>.

Bei einer Indexierung mit den spezifischen Deskriptoren der GND kann über die Inferenz zum Oberbegriff ›Malerei‹ eine transitive Hülle konstruiert werden, die eine Suchausweitung über die Hierarchie der Unterbegriffe für Malerei erlaubt<sup>123</sup>. Die grüne Fläche im folgenden Venn-Diagramm symbolisiert die transitive Hülle als mögliche Ergebnismenge bei diesem Szenario.

<sup>121</sup> Quelle: AAT Hierarchy Display, URL: <http://vocab.getty.edu/page/aat/300033799> (zitiert am 18.11.2018).

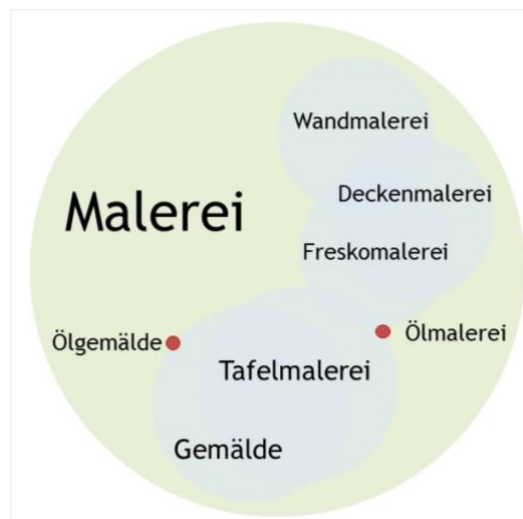
<sup>122</sup> ›Deckenmalerei‹ mit URI ›<http://d-nb.info/gnd/4011257-3>‹ ; ›Freskomalerei‹ mit URI ›<http://d-nb.info/gnd/4071463-9>‹ ; ›Gemälde‹ mit URI ›<http://d-nb.info/gnd/4122164-3>‹ ; ›Ölmalerei‹ mit URI ›<http://d-nb.info/gnd/4043238-5>‹ ; ›Malerei‹, mit URI ›<http://d-nb.info/gnd/4037220-0>‹ und ›Wandmalerei‹ mit URI ›<http://d-nb.info/gnd/4037220-0>‹.

<sup>123</sup> In den RDF-Repräsentationen der GND-Sachbegriffe ist die generische Hierarchierelation zum Oberbegriff durch die in der GND-Ontologie (GNDO) definierte Object Property »Broader term (generic)« (URI: <https://d-nb.info/standards/elementset/gnd#broaderTermGeneric>) ausgedrückt.



**Abbildung 2.18: Suchausweitung über GND-Hierarchierelationen am Beispiel »Malerei«**

Mit der Einbeziehung der Relationen zwischen den GND-Deskriptoren und den DDC- und GND-Sachgruppen ergibt sich ein besserer Recall für informationsorientierte Suchen. Über die Inferenz der Zugehörigkeit zur DDC-Sachgruppe ›750‹ (Malerei)<sup>124</sup> und GND-Sachgruppe ›13.3‹ (Malerei)<sup>125</sup> kann eine Begriffsgruppe gebildet werden, die alle mit den Begriffsgruppen verbundenen Deskriptoren in die Suche einbezieht. Im folgenden Venn-Diagramm ist diese Begriffsgruppe in der grünen Fläche dargestellt:



**Abbildung 2.19: Suchausweitung über DDC- und GND-Sachgruppen am Beispiel »Malerei«**

<sup>124</sup> Die Zugehörigkeit zu einer DDC-Sachgruppe wird anhand von vier in der GND definierten Properties ausgedrückt, die jeweils den Determiniertheitsgrad der in Beziehung stehenden DDC-Klasse bestimmen, z. B. »Related Dewey Decimal Classification with degree of determinacy 1« (URI: <https://d-nb.info/standards/elementset/gnd#relatedDdcWithDegreeOfDeterminacy1>). Die mit dem Determiniertheitsgrad 4 verknüpfte Klasse wird als Volltrefferklasse, die mit dem Determiniertheitsgrad 1 verknüpfte Klasse als Minimaltrefferklasse bezeichnet, vgl. Fachhochschule Köln (30. September 2010).

<sup>125</sup> Die Zugehörigkeit zu einer Sachgruppe der GND-Systematik wird in den RDF-Repräsentationen der GND-Sachbegriffe durch die GND-Property »GND subject category« (URI: <https://d-nb.info/standards/elementset/gnd#gndSubjectCategory>) ausgedrückt.

### 3 Untersuchungsmethode

Inwiefern die Lieferdaten der DDB tatsächlich als navigierbare Elemente die Suche im Portal unterstützen können, soll anhand einer Stichprobe analysiert werden. In diesem Kapitel wird erläutert, wie die Untersuchungsdaten für die Analyse erhoben und die Elemente aus den Metadatenschemata ausgewählt wurden. Es werden die Kriterien für die Stichprobenanalyse dargelegt, und das Vorgehen der Aufbereitung der Forschungsdaten wird erläutert.

#### 3.1 Erhebung einer Zufallsstichprobe aus den Lieferdaten

Zur Gewinnung des Untersuchungskorpus wurde eine randomisierte Stichprobe aus der Grundgesamtheit aller Datensätze in der DDB erhoben. Die Stichprobe sollte mindestens so groß sein, dass bezogen auf die Metadatenformate für die Datenlieferung an die DDB ein ähnliches prozentuales Verhältnis wie im Gesamtbestand vorliegt und je Format eine ausreichende Menge an Untersuchungsdatensätzen vorhanden ist. Ein Mengenverhältnis von einem Datensatz in der Stichprobe zu etwa 1.000 Datensätzen in der Grundgesamtheit schien für diesen Zweck angemessen.

Am 20.02.2017 wurde die öffentliche Programmierschnittstelle der DDB abgefragt und mittels einer Zufallsvariable in den Abfrageparametern eine randomisierte Auswahl von 20.000 Datensätzen aus dem DDB-Bestand erzeugt<sup>126</sup>. Als Grundgesamtheit dient die vom Application Programming Interface (API) der DDB<sup>127</sup> zum Abfragezeitpunkt gemeldete Anzahl von rund 20,5 Millionen Objektdatensätzen<sup>128</sup>. Zu jedem ermittelten Datensatz wurde der Identifikator des DDB-Objektes, der Namensraum des Metadatenformates, der Identifikator und Name der datengebenden Einrichtung sowie die zugehörige Sparte abgefragt. Diese Stichprobenmetadaten wurden zum Zweck der Datenanalyse in eine MySQL-Datenbank eingelesen.

Im nächsten Schritt wurden die erhobenen Daten in eine Excel-Arbeitsmappe überführt, in der anhand einer Merkmalsmatrix aus Metadatenformat und Datengeber insgesamt 201 Datensets gebildet und die für die Schnittmenge beider Merkmale zutreffende Anzahl an DDB-Objekten erfasst wurde<sup>129</sup>. Auf dieser Grundlage wurden die Daten im Hinblick auf mögliche Erhebungsfehler überprüft und Berechnungen zur Einschätzung der Aussagekraft der Stichprobe bezogen auf die Grundgesamtheit angestellt.

In der Stichprobe gibt es 16 fehlerhafte Datensätze. Drei Datensätze mussten aus dem Untersuchungskorpus entfernt werden, weil sie kein Kulturobjekt beschreiben. Für 13 Datensätze

---

<sup>126</sup> Das API der DDB wurde mit der URL »<https://api.deutsche-digitale-bibliothek.de/search>« und den Parametern »query=\*«, »rows=1« und »offset=offset« abgefragt. Bei der Variable »offset« handelt es sich um die Zufallszahl.

<sup>127</sup> Die Dokumentation des API der DDB findet sich auf der Website <https://pro.deutsche-digitale-bibliothek.de/node/530>.

<sup>128</sup> Die genaue Anzahl ist 20.530.483.

<sup>129</sup> siehe die Tabellendatei »ddb-stichprobe« in der Forschungsdatensammlung für diese Arbeit auf dem edoc-Server der Humboldt-Universität zu Berlin, permanenter Link: <https://doi.org/10.18452/19655>.



im MODS-Format ist der Datengeber nicht ermittelt worden. Diese Datensätze verblieben ohne Angaben zum Datengeber und zur Sparte im Untersuchungskorpus.

Die Stichprobe erfasste neun Metadatenformate, die nachfolgend in absteigender Reihenfolge nach der Häufigkeit ihres Vorkommens in der Stichprobe sortiert sind. Zuerst ist der offizielle Name des Schemas, in Klammern dahinter die in dieser Arbeit im Folgenden verwendete Abkürzung und hinter dem Komma der Uniform Resource Identifier des Namensraumes angegeben:

1. Encoded Archival Description (EAD), urn:isbn:1-931666-22-9
2. Metadata Object Description Standard (MODS), <http://www.loc.gov/mods/v3>
3. MARC 21 XML Schema (MARC), <http://www.loc.gov/MARC21/slim/>
4. Lightweight Information Describing Objects (LIDO), <http://www.lido-schema.org/>
5. Dublin Core Metadata Element Set (DCMES), <http://purl.org/dc/elements/1.1/>
6. DenkXweb, <http://www.rjm.de/denkxweb/denkxml/>
7. XML-Schema des Deutschen Filminstitutes (OAI\_DIF),  
[http://www.filmstandards.org/schemas/OAI/2.0/oai\\_dif/](http://www.filmstandards.org/schemas/OAI/2.0/oai_dif/)
8. XML-Schema des Social Science Open Access Repository (SSOAR),  
[http://www.ssoar.info/OAI/oai\\_ddb/](http://www.ssoar.info/OAI/oai_ddb/)
9. Europeana Data Model (EDM), <http://www.europeana.eu/schemas/edm/>

Folgende Tabelle zeigt für jedes Format, wie viele Datensätze von wie vielen Datengebern aufgeschlüsselt nach Spartenzugehörigkeit in der Stichprobe vorhanden sind.

**Tabelle 3.1: Übersicht der Datensätze in der Stichprobe nach Metadatenformat**

Format	Datensätze		Anzahl der Datengeber mit Spartenzugehörigkeit
	Anzahl	Anteil	
EAD	11.018	55,1 %	89 Datengeber der Sparte Archiv
MODS	3.165	15,8 %	25 Datengeber der Sparten: Bibliothek (20), Forschung (2), Archiv (1), Museum (1)
MARC	2.927	14,6 %	4 Datengeber der Sparte Bibliothek
LIDO	1.493	7,5 %	62 Datengeber der Sparten: Museum (50), Forschung (4), Archiv (3), Mediathek (3), Bibliothek (2)
DCMES	1.217	6,1 %	8 Datengeber der Sparten: Bibliothek (3), Forschung (2), Mediathek (2), Archiv (1)
DenkXWeb	71	0,36 %	4 Datengeber der Sparte Denkmalpflege
OAI_DIF	71	0,36 %	3 Datengeber der Sparten: Mediathek (2), Archiv (1)
SSOAR	32	0,16 %	1 Datengeber der Sparte Bibliothek
EDM	3	0,02 %	1 Datengeber der Sparte Bibliothek
<b>9 Formate</b>	<b>19.997</b>	<b>100 %</b>	<b>193 Datengeber<sup>130</sup> aus sechs Sparten</b>

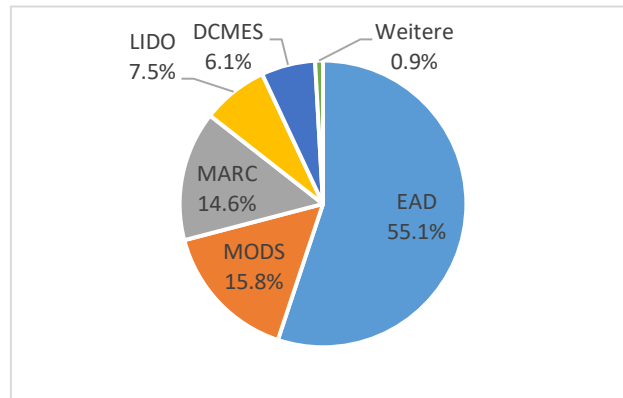
Um die Aussagekraft der Stichprobe einschätzen zu können, wurde der Anteil der Metadatenformate in der Stichprobe mit dem Anteil der Metadatenformate in der Grundgesamtheit verglichen. Weil es zum Ende der Erhebung (22.02.2017) nicht möglich war, mit vertretbarem Aufwand den Anteil der Metadatenformate in der Grundgesamtheit aller Datensätze in der DDB zu ermitteln, wurden die Vergleichszahlen einer von der Servicestelle DDB seit Mai 2018 geführten Excel-Tabelle zur Erfassung aller Datensets im DDB-Datenbestand herangezogen. In der für diese Arbeit vorliegenden Tabelle der »Provider-Datasets« ist der Datenbestand der DDB am 15.05.2018 mit rund 24 Millionen Datensätzen<sup>131</sup> erfasst. Aus dieser Tabelle erschließt sich, dass das Zufallsverfahren fast alle Lieferformate der DDB erhoben hat. Die einzige Ausnahme bildet index.meta, ein Datenformat eines Datengebers der Sparte »Forschung«, das in der Grundgesamtheit aber einen Anteil von nur 0,01 Prozent ausmacht.

Aus den nachfolgenden Diagrammen ist ersichtlich, dass sich die Metadatenformate in der Stichprobe und in der Grundgesamtheit prozentual ähnlich verteilen (siehe Abbildung 3.1 mit 19.997 Datensätzen in der Stichprobe und Abbildung 3.2 mit 24.158.148 Datensätzen in der

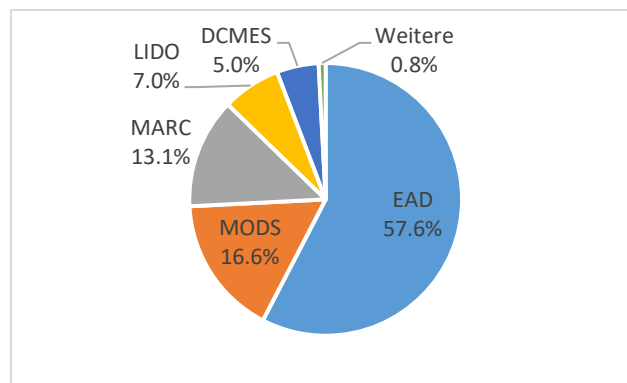
<sup>130</sup> Die Summe der Datengeber nach den Formaten (197) ist größer als die tatsächliche Anzahl der Datengeber (193), weil es Datengeber gibt, die verschiedene Bestände in verschiedenen Metadatenformaten an die DDB liefern. Beispielsweise lieferte das Thüringische Hauptstaatsarchiv Weimar einen Bestand im Format EAD und einen Bestand im Format MODS an die DDB. In der Grundgesamtheit lagen zum Erhebungszeitpunkt am 20.02.2017 Datensätze von 284 Datengebern vor.

<sup>131</sup> Siehe die Tabellendatei »ddb-provider-datasets« in der Forschungsdatensammlung für diese Arbeit auf dem edoc-Server der Humboldt-Universität zu Berlin, permanenter Link: <https://doi.org/10.18452/19655>.

Grundgesamtheit im Vergleich). Die Metadatenformate mit einem Anteil von unter einem Prozent wurden in beiden Diagrammen zur besseren Übersichtlichkeit in der Gruppe »Weitere« zusammengefasst. Die größte Verschiebung stellt der Anteil der Metadaten im EAD-Format dar, der in der Stichprobe um 2,5 Prozent geringer ist als in der Grundgesamtheit. Diese Abweichung kann dadurch erklärt werden, dass im Zeitraum von Februar 2017 bis Mai 2018 ein überproportional höherer Zuwachs an Datensätzen aus der Sparte »Archiv« erfolgte<sup>132</sup>.



**Abbildung 3.1: Verteilung der Metadatenformate in der Stichprobe (22.02.2017)**



**Abbildung 3.2: Verteilung der Metadatenformate in der Grundgesamtheit (15.05.2018)**

Die Metadatenformate mit einem Anteil von unter einem Prozent werden für die weitere Untersuchung nicht berücksichtigt, weil zu wenige Datensätze in diesen Formaten vorliegen, um daraus aussagekräftige Ergebnisse zu gewinnen. Die Untersuchung bezieht sich daher auf das archivische Metadatenformat EAD, die bibliothekarischen Formate MODS und MARC, das Museumsformat LIDO und das übergreifende Elementset des Dublin-Core-Standards DCMES. Mit diesen fünf Formaten können Aussagen zu etwa 99 Prozent aller Datensätze in der DDB getroffen werden. Es verbleiben somit 19.874 Datensätze im Untersuchungskorpus. Die Metadatenformate werden zu Beginn der Präsentation der Ergebnisse der Stichprobenanalyse in Kapitel 4 jeweils kurz vorgestellt.

<sup>132</sup> siehe Geschäftsführung der Deutschen Digitalen Bibliothek (2. Mai 2018, Kapitel 4: Entwicklung des Datenbestandes).

## 3.2 Auswahl der Elemente aus den Metadatenschemata

Ausgangspunkt für die Untersuchung bildet eine Konkordanz von Elementen aus den Metadatenschemata zu den vorgeschlagenen Filtern für die DDB. Bei der Auswahl der Elemente und Attribute wurde folgenden Fragen nachgegangen:

1. Definiert das Metadatenschema ein Element, dessen Inhalt für den jeweiligen vorgeschlagenen Filter ausgewertet werden kann?
2. Gibt das Metadatenschema die Anwendung des Elementes verpflichtend vor?
3. Unterscheiden sich die Bedeutungen der Metadatenelemente, die einem der vorgeschlagenen Filter zugeordnet werden können, voneinander?
4. Wie sollen Indexterme aus kontrollierten Vokabularen in den Elementinstanzen ausgewiesen werden?
5. Machen die Schemata und Anwendungsprofile Vorgaben oder Empfehlungen zur Verwendung bestimmter kontrollierter Vokabulare?

Zur Klärung der Bedeutung der Metadatenelemente wurden Definitionen, Anwendungshinweise und Konkordanzen aus den Spezifikationen und Anwendungsleitfäden der Schemata und Anwendungsprofile zusammengetragen. Als zusätzliches Hilfsmittel diente der »Metadata Standards Crosswalk« des Getty Research Institute<sup>133</sup> mit seinen Konkordanzen zwischen den Elementen verschiedener Metadatenstrukturstandards des Kulturerbe-Bereiches.

Die zu jedem Metadatenelement ermittelten Informationen wurden in tabellarischer Form aufbereitet. Diese Tabellen sind geordnet nach Metadatenschema im »Anhang A: Zusammenstellung der untersuchten Metadatenelemente« zu finden. Die Bezeichnungen der Elemente sind darin in englischer und, wenn eine Übersetzung vorliegt, auch in deutscher Sprache angegeben. Die deutschsprachige Variante wird im Fließtext dieser Arbeit bevorzugt verwendet. Der Verpflichtungsgrad und die Wiederholbarkeit der Elemente sind in den Tabellen einheitlich in der Chen-Notation ausgedrückt. Unterelemente und Attribute sind nur dann berücksichtigt worden, wenn sie für die Spezifizierung von Indextermen und für Quellennachweise vorgesehen sind.

### 3.2.1 Filter »Sachkategorie«

Ein Element für den vorgeschlagenen Filter »Sachkategorie« definieren die bibliothekarischen Standards MARC und MODS, das Museumsformat LIDO und das übergreifende Elementset des Dublin-Core-Standards (DCMES). In keinem dieser Formate ist das Element verpflichtend.

**Tabelle 3.2: Metadatenelemente für den vorgeschlagenen Filter »Sachkategorie«**

MARC	MODS	EAD	LIDO	DCMES
082, 083, 084	<classification>	-	<classification>	<subject>

<sup>133</sup> vgl. The J. Paul Getty Trust (24. Oktober 2017).

Die Bedeutung des Elements unterscheidet sich abhängig von der Sparte: MARC und MODS definieren das Element im Sinne der bibliothekarischen Tradition als Element der klassifikatorischen Inhaltserschließung, »die mit Hilfe von Notationen Inhalte wiedergibt, anzeigt, nach sachlichen Kriterien ordnet und suchfähig macht« (Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung, März 2017, S. 35). Die Vorgaben für »Classification« im Museumsformat LIDO basieren auf dem XML-Schema CDWA Lite der »Categories for the Description of Works of Art«<sup>134</sup>. Die Inhalte des LIDO-Elements »Classification« sollen Objekte aufgrund gleicher Merkmale zusammenstellen und gruppieren. Diese Merkmale können zum Beispiel die Zugehörigkeit zu einer Disziplin wie »Decorative Arts« oder zu einer Objektgattung wie »Sculpture« sein. Der Anwendungshinweis für das LIDO-Element (»How to record«) und die im Leitfaden »Implementing LIDO« vorgeschlagene Klassifizierung ziehen weitere unterscheidende Merkmale heran, wie Stil, Form, Alter, Geschlecht und Phase<sup>135</sup>. Im übergreifenden DCMES können Notationen im Element für die Beschreibung des Themas der Ressource angegeben werden.

Alle Schemata empfehlen die Verwendung von kontrolliertem Vokabular. Die bibliothekarischen Standards MARC und MODS geben die zur Anwendung empfohlenen Klassifikationssysteme und Systematiken in einer verbindlichen Liste vor. Die Liste der »Classification Scheme Source Codes« des Network Development and MARC Standards Office der Library of Congress enthält gleichzeitig Codes zum Nachweis der Quelle der Notationen in den bibliografischen Metadaten<sup>136</sup>. Das LIDO-Schema bietet ebenfalls ein Attribut, um die Quelle auszuzeichnen und empfiehlt generell, dass Indexterme bevorzugt aus einem publizierten Vokabular stammen sollen. Empfehlungen für bestimmte Vokabulare sind in den LIDO-Anwendungsprofilen »DDB-LIDO«<sup>137</sup> und »Gemeinsamer Feldkatalog Graphischer Sammlungen« zu finden<sup>138</sup>. Im DCMES kann die Quelle des Indexterms nicht explizit ausgewiesen werden.

Nur die Schemata MODS und LIDO weisen ein Element oder Attribut aus, in dem der Indexterm für die »Sachkategorie« durch einen URI referenziert werden kann.

### 3.2.2 Filter »Themenschlagwort«

Alle Metadatenschemata definieren ein Element für den vorgeschlagenen Filter »Themenschlagwort«. Das Element ist in keinem Format verpflichtend.

**Tabelle 3.3: Metadatenelemente für den vorgeschlagenen Filter »Themenschlagwort«**

MARC	MODS	EAD	LIDO	DCMES
650	<topic>	<subject>	<subjectConcept>	<subject>

<sup>134</sup> vgl. J. Paul Getty Trust und ARTstor (17. Juli 2006, S. 23).

<sup>135</sup> vgl. McKenna und Stein (2015, S. 7 und 39 ff.).

<sup>136</sup> vgl. Library of Congress - Network Development and MARC Standards Office (13. April 2018). Ergänzende Vorgaben zur Codierung sind in den Hinweisen der MARC- und MODS-Anwendungsprofile zu finden.

<sup>137</sup> vgl. Fachstelle Museum der Deutschen Digitalen Bibliothek im Institut für Museumsforschung (2014).

<sup>138</sup> vgl. Fichtl et al. (2015).

Das Element dient in allen Schemata dazu, das Thema eines Kulturobjektes mittels geeigneter Indexterme zu beschreiben. In MARC und MODS gibt das Element im Sinne der bibliothekarischen Erschließungstradition den Inhalt des Dokumentes durch Sachschlagwörter auf spezifische Weise wieder. Das EAD-Element »Gegenstand« ist für Begriffe vorgesehen, die das von der Archivalie behandelte Thema bezeichnen. Im entsprechenden LIDO-Element sollen Sachbegriffe referenziert werden, die das Thema eines Museumsobjektes beschreiben. Dabei kann es sich zum Beispiel um das Motiv eines Werkes der bildenden Kunst handeln. Das DCMES-Element zur Beschreibung des Themas der Ressource kann nicht nur Schlagwörter, sondern auch Stichwörter und Notationen enthalten.

Alle Schemata empfehlen die Verwendung von kontrolliertem Vokabular. Die Bezugsquellen für die Indexterme in den bibliothekarischen Standards MARC und MODS finden sich in der Liste der »Subject Heading and Term Source Codes« des Network Development and MARC Standards Office der Library of Congress<sup>139</sup>. Auch die Schemata EAD und LIDO ermöglichen es, die Quelle des Indexterms in einem Attribut zu kennzeichnen. In den LIDO-Anwendungsprofilen sind Vorgaben für bestimmte kontrollierte Vokabulare zu finden.

Alle Schemata außer DCMES weisen ein Element oder Attribut aus, in dem der Indexterm für das Themenschlagwort referenziert werden kann: entweder durch die Identifikationsnummer des Normdatensatzes oder durch den URI des Begriffes im verwendeten kontrollierten Vokabular.

### 3.2.3 Filter »Objektgattung«

Alle Metadatenschemata definieren ein Element für den vorgeschlagenen Filter »Objektgattung«. Das Element ist nur im Austauschformat LIDO verpflichtend.

**Tabelle 3.4: Metadatenelemente für den vorgeschlagenen Filter »Objektgattung«**

MARC	MODS	EAD	LIDO	DCMES
655	<genre>	<genreform>	<objectWorkType>	<type>

Das Element dient in allen Schemata dazu, den generischen Typ des beschriebenen Kulturobjektes auszuweisen. Die bibliothekarischen Standards MARC und MODS definieren das Element als ein die Form kennzeichnendes Schlagwort, das den Typ des Dokumentes nach Form oder Inhalt spezifiziert. Das EAD-Element »Genre- und Formbegriffe« für die Kennzeichnung des Archivalientyps entspricht dieser Definition weitgehend. Im LIDO-Format beschreibt das Element den Typ des Objektes auf spezifischer Ebene und im übergreifenden Elementset des Dublin-Core-Standards die Art oder Gattung der Ressource.

Alle Schemata empfehlen die Verwendung eines kontrollierten Vokabulars. Bezugsquellen für Genre- und Forms Schlagwörter in MARC und MODS sind in der Liste der »Genre/Form Code and

<sup>139</sup> vgl. Library of Congress - Network Development and MARC Standards Office (7. September 2018).

Term Source Codes« des Network Development and MARC Standards Office der Library of Congress der Library of Congress verzeichnet<sup>140</sup>. Das EAD(DDB)-Profil für Datenlieferungen an die DDB stellt seinen Anwendern für das Attribut des normierten Archivalientyps eine Auswahlliste mit zwölf Werten bereit<sup>141</sup>. Auch die LIDO-Anwendungsprofile machen Vorgaben für kontrollierte Vokabulare. Das DCMES benennt als mögliche Bezugsquelle für den Indexterm zur Typisierung der Ressource das »DCMI Type Vocabulary«<sup>142</sup>.

Alle Schemata außer DCMES weisen ein Element oder Attribut aus, in dem der Indexterm für die Objektgattung referenziert werden kann: entweder durch die Identifikationsnummer des Normdatensatzes oder durch den URI des Begriffes im verwendeten kontrollierten Vokabular.

### 3.3 Untersuchungskriterien

Im Rahmen der Stichprobenanalyse soll untersucht werden, ob und, wenn ja, inwieweit die objektbeschreibenden Metadaten der DDB die in Kapitel 2.3 und 2.4 beschriebenen Voraussetzungen für eine Suchunterstützung mittels navigierbarer Informationsfilter erfüllen. Um diese Forschungsfrage adäquat beantworten zu können, werden im Folgenden quantitative und qualitative Untersuchungskriterien hergeleitet.

#### **(1) Datensätze mit Element für die vorgeschlagenen Filter (quantitativ)**

Als Voraussetzung wurde festgehalten, dass in jedem Metadatensatz mindestens ein Element für jeden vorgeschlagenen Filter vorhanden sein muss. Bei der Erstellung der Konkordanz wurde bereits festgestellt, dass nicht alle Schemata ein Element für jeden Filter definieren<sup>143</sup> und bis auf eine Ausnahme kein Element in den Lieferformaten verpflichtend ist<sup>144</sup>. Das erste Untersuchungskriterium gilt daher der Beantwortung der Frage:

- In wie vielen Datensätzen ist das Element für den vorgeschlagenen DDB-Informationsfilter mindestens einmal vorhanden und enthält einen Wert für den Indexterm?

#### **(2) Auslegung der Definition des Metadatenelementes (qualitativ)**

Damit die Elemente in den Lieferdaten zuverlässig für die vorgeschlagenen Filter ausgewertet werden können, müssen sie von den Datengebern spezifikationskonform angewendet worden sein. Für jeden vorgeschlagenen Filter soll daher die Frage beantwortet werden:

- Inwiefern entsprechen die Inhalte des Elementes für den jeweiligen Informationsfilter der Definition des Elementes im entsprechenden Metadatenschema?

---

<sup>140</sup> vgl. Library of Congress - Network Development and MARC Standards Office (8. Juni 2018). Weitere Listen und Codierungen sind in den Hinweisen der Anwendungsprofile aufgeführt.

<sup>141</sup> vgl. Krauth (8. Januar 2013, Tabelle »Findbuch«); Landesarchiv Baden-Württemberg (2013, S. 6, Fußnote 28)

<sup>142</sup> vgl. DCMI Usage Board (14. Juni 2012, Section 7: DCMI Type Vocabulary).

<sup>143</sup> Das EAD-Schema besitzt kein Element für den Filter »Sachkategorie«. Das DCMES-Elementset definiert nur ein Element »Thema«, das nicht differenziert auf die Filter »Themenschlagwort« und »Sachkategorie« abgebildet werden kann.

<sup>144</sup> Das einzige Pflichtelement dieser Untersuchung ist »Object/Work Type« im LIDO-Schema.

### **(3) Indexterme aus kontrolliertem Vokabular (quantitativ und qualitativ)**

Mit diesem Kriterium soll die tatsächliche Ausgangslage bezüglich der Indexierung mit kontrolliertem Vokabular in der DDB erhoben werden. Für jeden vorgeschlagenen Filter sollen daher folgende Fragen beantwortet werden:

1. Wie viele der Indexterme stammen aus einem kontrollierten Vokabular?
  - a) In welchen Vokabularen sind diese Indexterme verzeichnet?
  - b) Wie viele Indexterme liegen je Vokabular vor?
  - c) Wie viele Indexterme sind jeweils durch einen URI eindeutig referenziert?
2. Wie viele der Indexterme sind ungebundene Schlagwörter oder Deskriptoren aus hauseigenen Vokabularen?
3. Inwiefern eignen sich die von den Datengebern verwendeten kontrollierten Vokabulare als Navigationsinstrumente für die vorgeschlagenen Filter?

Im Rahmen einer Datenauswertung wird jeweils die Anzahl aller Elementwerte und davon wiederum die der kontrollierten und unterscheidbaren Werte ermittelt. Als kontrolliert gezählt werden Indexterme, (a) deren Quelle durch eine identifizierbare Angabe belegt ist oder (b) die durch einen URI eindeutig mit einem Vokabularbegriff verbunden sind.

Indexterme ohne identifizierbare Quelle zählen in dieser Untersuchung nicht zu den kontrollierten Werten. In diese Gruppe fallen sowohl ungebundene Schlagwörter als auch Deskriptoren aus hauseigenen Vokabularen, die aufgrund der Erhebungsmethode nicht voneinander unterschieden werden können.

Als unterscheidbar zählen jeweils alle Indexterme mit identischen Zeichenketten. Dieser Indikator dient der Einschätzung der Diskriminanz und Indexierungsspezifität im jeweiligen Erschließungselement.

## **3.4 Aufbereitung der Forschungsdaten**

Für die Datenauswertung wurden die Untersuchungsdaten aus der MySQL-Datenbank im Dateiformat CSV<sup>145</sup> ausgelesen und in eine Tabellenkalkulation eingelesen. Für jedes Metadatenformat gibt es eine Haupttabelle, in der alle Element- und Attributwerte aus einem Datensatz in einer Zeile angeordnet sind: »marc\_stichprobe«, »mods\_stichprobe«, »ead\_stichprobe«, »lido\_stichprobe« und »dc\_stichprobe«.

Um die Bezüge einzelner Element- und Attributinstanzen zueinander darzustellen, wurden die Werte aus mehrfach verwendeten Elementen mit einem Verkettungszeichen » | « voneinander abgegrenzt und nicht vorhandene Werte mit einem Bindestrich » - « gekennzeichnet.

Folgendes Beispiel illustriert, wie Instanzen des Elementes »Classification« in der Extensible Markup Language (XML) im MODS-Format abgebildet sind:

---

<sup>145</sup> comma-separated values



```

<mods:classification authority="gdz">RusDML</mods:classification>
<mods:classification authority="gdz">Mathematica</mods:classification>
<mods:classification>513.83</mods:classification>

```

In der Tabelle »mods\_stichprobe« sind diese Instanzen folgendermaßen dargestellt:

**Tabelle 3.5: Elementinstanzen in der Tabelle »mods\_stichprobe«**

DDB-Identifikator	classification_authority	classification
BYJJOVSP5GDBAE2552MLYVDV7BMWOIMF	gdz   gdz   -	RusDML   Mathematica   513.83

Anhand dieser Tabellen wurde die Anzahl der Datensätze mit dem jeweiligen Element und Attribut berechnet. Im nächsten Schritt wurden die Werte aus den Haupttabellen in einzelne Elementinstanzen aufgeteilt. Für jedes Metadatenelement gibt es eine Tabelle, in der jede Instanz eines Elementes in einer Zeile dargestellt ist:

**Tabelle 3.6: Elementinstanzen in der Tabelle »mods\_classification«**

DDB-Identifikator	classification_authority	classification
BYJJOVSP5GDBAE2552MLYVDV7BMWOIMF	gdz	RusDML
BYJJOVSP5GDBAE2552MLYVDV7BMWOIMF	gdz	Mathematica
BYJJOVSP5GDBAE2552MLYVDV7BMWOIMF	-	513.83

Die Elementtabellen bildeten die Grundlage für die Berechnung der Anzahl der Werte in den Elementen und Attributen.

Die Forschungsdaten befinden sich auf dem edoc-Server der Humboldt-Universität zu Berlin und sind dort über den Digital Object Identifier <https://doi.org/10.18452/19655> zugänglich.

## 4 Stichprobenanalyse nach Metadatenformat

Dieses Kapitel präsentiert die Ergebnisse der Stichprobenanalyse für die einzelnen Metadatenformate in der Reihenfolge der Sparten, in denen die Formate hauptsächlich verwendet werden: Bibliothek, Archiv und Museum (BAM) und weitere. Eine kurze Vorstellung des jeweiligen Schemas und dessen Besonderheiten in Bezug auf die untersuchten Metadatenelemente erfolgt zu Beginn eines jeden Formatkapitels. Die Ergebnisse der Datenauswertung sind für die einzelnen vorgeschlagenen Informationsfilter in Form von Tabellen dargestellt, worin die Zeile des jeweiligen Elementes für den Indexterm grau unterlegt ist. Im darauffolgenden Abschnitt ist jeweils dargelegt, inwiefern die Elementwerte der Definition des Elementes im Metadatenstandard entsprechen. Hierfür wurde stichprobenartig die Bedeutung der häufigsten Elementwerte mit der Definition im Standard verglichen. Die von den Datengebern für die Indexierung verwendeten Vokabulare werden danach in einem separaten Abschnitt behandelt.

### 4.1 MACHine-Readable Cataloging (MARC)

Das XML-Schema des vom Network Development and MARC Standards Office der Library of Congress entwickelten Metadatenstandards MARC 21 Bibliographic wird international für den Austausch von bibliografischen Metadaten genutzt. Für deutschsprachige Anwender der MARC-Standards ist die Arbeitsstelle für Standardisierung an der Deutschen Nationalbibliothek (DNB) Ansprechpartnerin.

Die Stichprobe enthält 2.927 Datensätze im MARCXML-Format von vier Datengebern der Sparte Bibliothek. Der Anteil der MARC-Datensätze beträgt 14,6 Prozent in der Stichprobe und 13,6 Prozent in der Grundgesamtheit aller DDB-Datensätze. Weil die DNB die Datengeberin mit den meisten Datensätzen in der Stichprobe ist, wurde bei der Interpretation der Elementinhalte primär die »Feldbeschreibung der Titeldaten der Deutschen Nationalbibliothek und Zeitschriftendatenbank im Format MARC 21« berücksichtigt<sup>146</sup>.

Die MARCXML-Elemente werden in den Tabellen dieses Kapitels in der Schreibweise von MARC 21 wiedergegeben. Das Format bietet die Möglichkeit, kontrollierte Indexterme auszuweisen. Das Unterfeld \$0 ist für die Identifikationsnummer des Normdatensatzes und das Unterfeld \$2 für die Angabe der Quelle des Indexterms vorgesehen.

---

<sup>146</sup> vgl. Deutsche Nationalbibliothek (16. Juni 2014).

### 4.1.1 Sachkategorie

**Tabelle 4.1: Datenauswertung der MARC-Elemente für Notationen**

MARC-Feld und Unterfeld	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Feld/Unterfeld	mit Werten im Feld/Unterfeld	insgesamt	kontrolliert	unterscheidbar
082 \$a	1.160	1.160	1.298	1.254	655
082 \$2	1.128	1.128	1.254		9
083 \$a	565	565	784	784	132
083 \$2	565	565	784		3
084 \$a	1.874	1.874	4.574	4.549	577
084 \$2	1.874	1.874	4.573		10
Zusammenfassung	Datensätze mit mindestens einem MARC-Element für Notationen: 65,0 Prozent		Indexterme aus kontrolliertem Vokabular: 98,5 Prozent, davon Indexterme mit URI: 0 Prozent		

#### Auslegung der Definition des Metadatenelementes

Die MARC-Elemente für Notationen werden von den Datengebern standardkonform ausgelegt. Die Elemente für die Dewey-Dezimalklassifikation (DDC) enthalten ihrer Definition zufolge Notationen aus der DDC. Verwendet werden vielfach Notationen, die die Klassen der zweiten Hierarchieebene der DDC bezeichnen: Die häufigsten Werte in diesen MARC-Elementen sind ›610‹ (Anzahl: 369; Klassenbezeichnung: ›Medizin und Gesundheit‹), ›330‹ (Anzahl: 193; Klassenbezeichnung: ›Wirtschaft‹) und ›340‹ (Anzahl: 188; Klassenbezeichnung: ›Recht‹). Weitere Notationen aus anderen Klassifikationssystemen oder Systematiken sind im dafür jeweils vorgesehenen Element enthalten.

#### Indexterme aus kontrolliertem Vokabular

Nahezu alle Werte in den MARC-Elementen für Notationen stammen aus Klassifikationssystemen und Systematiken, die im Bibliotheksbereich verbreitet sind (93 Prozent). Die Quelle der Notation ist im entsprechenden Unterfeld zuverlässig angegeben. Am häufigsten verwendet wird die DDC, und zwar in verschiedenen Ausgaben und Varianten. Die DDC-Sachgruppen der Deutschen Nationalbibliografien werden vor allem von der DNB genutzt. Diese Sachgruppen ordnen die Dokumente in eine grobe fachliche Gliederung mit circa 100 auf der DDC basierenden Klassen ein<sup>147</sup>. Der Anteil der Notationen aus der Regensburger Verbundklassifikation<sup>148</sup> ist mit zwei Prozent auffallend gering, weil dieses Sacherschließungsinstrument vor allem von universitären und

<sup>147</sup> siehe dazu Kapitel 2.4.3 dieser Arbeit.

<sup>148</sup> siehe Universitätsbibliothek Regensburg (2018).

wissenschaftlichen Bibliotheken genutzt wird, die unter den MARC-Datengebern der DDB nicht vertreten sind.

**Tabelle 4.2: Kontrollierte Vokabulare in den MARC-Elementen für Notationen**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
DDC-Sachgruppen der Deutschen Nationalbibliografien	sdbn	4.286	64,4 %	0
Dewey-Dezimalklassifikation	19, 20, bio01, 22/ger, 22/eng, 22//ger	945	14,2 %	0
GND-Sachgruppen	sswd	582	8,7 %	0
Sondersammelgebietsnummern <sup>149</sup>	ssgn	235	3,5 %	0
Regensburger Verbundklassifikation	rvk	134	2,0 %	0

#### 4.1.2 Themenschlagwort

**Tabelle 4.3: Datenauswertung des MARC-Elementes »Sachschlagwort«**

MARC-Feld und Unterfeld	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Feld/Unterfeld	mit Werten im Feld/Unterfeld	insgesamt	kontrolliert	unterscheidbar
650 \$a	932	932	2.561	2.201	1.984
650 \$0	765	765	2.039	2.039	1.603
650 \$2	907	907	2.508		5
Zusammenfassung	Datensätze mit mindestens einem MARC-Element »Sachschlagwort«.: 31,8 Prozent		Indexterme aus kontrolliertem Vokabular: 85,9 Prozent, davon Indexterme mit URI: 85,9 Prozent		

#### Auslegung der Definition des Metadatenelementes

Das MARC-Element »Sachschlagwort« enthält überwiegend Bezeichnungen für Allgemeinbegriffe, die den Inhalt der Dokumente wiedergeben: Die häufigsten Werte sind »Deutsch« (Anzahl: 22), »Literatur« (Anzahl: 21) und »Unternehmen« (Anzahl: 17). Das Element wird, wenn es vorhanden ist, von den Datengebern meistens standardkonform verwendet. Einige wenige Fehlinterpretationen betreffen Elementwerte, die den Dokumenttyp ausweisen, zum Beispiel »Zeitschrift« (Anzahl: 8) und »Online-Ressource« (Anzahl: 6).

<sup>149</sup> Dabei handelt es sich um eine Nummerierung der Fachgebiete im Sondersammelgebietsplan der Deutschen Forschungsgemeinschaft (DFG), siehe Deutsche Forschungsgemeinschaft (2015, Anhang).

## Indexterme aus kontrolliertem Vokabular

Die meisten Sachschlagwörter sind kontrollierte Werte aus der Gemeinsamen Normdatei (GND) (86 Prozent). Der URI für den GND-Sachbegriff ist anhand der Identifikationsnummer des Normdatensatzes konstruierbar. Die entsprechenden Elementwerte enthalten den International Standard Identifier for Libraries and Related Organisations (ISIL) für die GND<sup>150</sup>, anhand dessen der Namensraum für den GND-URI gebildet werden kann, zum Beispiel von der Identifikationsnummer ›(DE-588)4045791-6‹ zum URI ›http://d-nb.info/gnd/4045791-6‹. Die entsprechend ausgewiesenen Elementwerte zählen in dieser Datenauswertung daher als Indexterme mit URI.

**Tabelle 4.4: Kontrolliertes Vokabular im MARC-Element »Sachschlagwort«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Gemeinsame Normdatei	gnd, swd, ISIL (DE-588) in § 0	2.201	85,9 %	2.201

### 4.1.3 Objektgattung

**Tabelle 4.5: Datenauswertung des MARC-Elementes »Genre/Formschlagwort«**

MARC-Feld und Unterfeld	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Feld/Unterfeld	mit Werten im Feld/Unterfeld	insgesamt	kontrolliert	unterscheidbar
655 \$a	516	516	654	622	111
655 \$0	71	71	82	82	29
655 \$2	497	497	623		5
Zusammenfassung	Datensätze mit mindestens einem MARC-Element »Genre-/Formschlagwort«: 17,6 Prozent		Indexterme aus kontrolliertem Vokabular: 95,1 Prozent, davon Indexterme mit URI: 12,5 Prozent		

## Auslegung der Definition des Metadatenelementes

Das MARC-Element »Genre/Formschlagwort« wird, wenn es vorhanden ist, von den Datengebern überwiegend standardkonform verwendet. Die häufigsten Werte sind ›Aufsatzsammlung‹ (Anzahl: 108), ›Kongress‹ (Anzahl: 64) und ›Online-Publikation‹ (Anzahl: 39). Diese Schlagwörter kennzeichnen den Dokumenttyp entweder nach der Art seines Inhaltes (z. B. ›Aufsatzsammlung‹) oder nach seiner Form (z. B. ›Online-Publikation‹). Einige Indexterme weisen jedoch keinen Bezug zur Art des Dokumentes auf, wie zum Beispiel ›Kongress‹.

## Indexterme aus kontrolliertem Vokabular

<sup>150</sup> Der URI <http://d.zdb-services.de/resource/organisations/DE-588> für den ISIL ›DE-588‹ löst im Webbrowser auf den Eintrag für die Gemeinsame Normdatei im Verzeichnis der Deutschen ISIL-Agentur und Sigelstelle auf.

Das meistgenutzte Vokabular für die in den MARC-Daten ausgewiesenen Genre- und Forms Schlagwörter ist die GND (84 Prozent). Nur für einen kleinen Anteil dieser Indexterme ist jedoch der GND-URI für den Sachbegriff angegeben. Die meisten Indexterme mit der GND als Quellenangabe sind nicht durch eine Identifikationsnummer mit dem Normdatensatz verbunden. Des Weiteren finden sich Schlagwortlisten mit Gattungsbegriffen unter den Quellen für die Genre- und Forms Schlagwörter. Beide Schlagwortlisten sind mit dem Code ›gatbeg‹ gekennzeichnet. Voneinander unterscheidbar sind die Indexterme aus diesen Listen nur, weil die Gattungsbegriffe der DNB mit einer Identifikationsnummer gekennzeichnet sind, die den ISIL der DNB als zusätzliches kennzeichnendes Merkmal enthalten.

**Tabelle 4.6: Kontrollierte Vokabulare im MARC-Element »Genre-/Formschlagwort«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Gemeinsame Normdatei	gnd, gnd-content, gnd-music	539	83,9 %	82
Gattungs- und Sachbegriffe der Arbeitsgemeinschaft Alter Drucke	gatbeg	52	8 %	0
Liste der in der Deutschen Nationalbibliothek verwendeten Gattungsbegriffe <sup>151</sup>	gatbeg, DE-101 in § 0	21	3,2 %	0

## 4.2 Metadata Object Description Schema (MODS)

Der vom Network Development and MARC Standards Office der Library of Congress entwickelte Metadatenstandard MODS wird zur bibliografischen Beschreibung von digitalisierten Medien verwendet. MODS-basierte Datenlieferungen an die DDB entstammen zum großen Teil Projekten zur Digitalisierung von Druckwerken, die von der Deutschen Forschungsgemeinschaft (DFG) gefördert wurden. Durch die gemeinsame Verwendung des »MODS-Anwendungsprofils für digitalisierte Medien« sollen die Metadaten soweit vereinheitlicht werden, dass sie in übergreifenden Anwendungen wie dem DFG-Viewer<sup>152</sup>, dem Zentralen Verzeichnis Digitalisierter Drucke (ZVDD)<sup>153</sup> und der DDB durchsuchbar gemacht werden können<sup>154</sup>.

Die Stichprobe enthält 3.165 Datensätze im MODS-Format von 25 Datengebern aus vier Sparten: Bibliothek (20), Forschung (2), Archiv (1), Museum (1). Der Anteil der MODS-Datensätze beträgt 15,8 Prozent in der Stichprobe und 16,6 Prozent in der Grundgesamtheit aller DDB-Datensätze.

<sup>151</sup> siehe Deutsche Nationalbibliothek (25. Juli 2012).

<sup>152</sup> Dabei handelt es sich um eine Referenzimplementierung für Digitalisierungsstandards der Deutschen Forschungsgemeinschaft (DFG), siehe <http://dfg-viewer.de/>.

<sup>153</sup> siehe <http://www.zvdd.de/>.

<sup>154</sup> vgl. Rühle, Jahnke und Kühle (2017, S. 6).

Das MODS-Format bietet die Möglichkeit, kontrollierte Indexterme in Form eines URI auszuweisen, die allerdings sehr selten wahrgenommen wird. Wenn kontrollierte Werte vorhanden sind, wird das Vokabular durch einen Code im Authority-Attribut angegeben.

#### 4.2.1 Sachkategorie

**Tabelle 4.7: Datenauswertung des MODS-Elementes »Klassifikation«**

MODS-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/Attribut	insgesamt	kontrolliert	unterscheidbar
<classification>	2.132	2.132	4.060	148	476
valueURI	0		0		
authority	2.105	2.105	4.002	3	29
authorityURI	0		0		
Zusammenfassung	Datensätze mit mindestens einem MODS-Element »Klassifikation«: 67,4 Prozent		Indexterme aus kontrolliertem Vokabular: 3,2 Prozent, davon Indexterme mit URI: 0 Prozent		

#### Auslegung der Definition des Metadatenelementes

Das Element »Klassifikation« wird entgegen seiner Definition im MODS-Standard überwiegend nicht für die Vergabe von Notationen verwendet, wie bereits die häufigsten Indexterme ›paedagogischezeitschriften‹ (Anzahl: 444), ›VD18digital‹ (Anzahl: 255) und ›370.education‹ (Anzahl: 237) aufzeigen. Die meisten Datengeber nutzen dieses Element, um ihre Druckwerke verschiedenen Sammlungen zuzuweisen. Diese Schlussfolgerung ergibt sich aus den »Mindestanforderungen für die Bereitstellungssysteme von Digitalisaten« in den DFG-Praxisregeln. Demnach soll den Nutzern dieser Systeme eine Möglichkeit geboten werden, »in vorab definierten Sammlungen, Sammlungsteilen oder Beständen strukturiert zu navigieren (browsing)« (Deutsche Forschungsgemeinschaft, 12/2016, S. 44)<sup>155</sup>. Das MODS-Element »Klassifikation« dient der Ermöglichung derartiger Recherchefunktionen. Die Elementwerte in der Stichprobe entsprechen vielfach den Sammlungsnamen der navigierbaren Zugänge in den lokalen Präsentationssystemen der Datengeber. Beispielsweise entspricht der Elementwert ›Hamburgensien: Zeitschriften‹ (Anzahl: 164) einer navigierbaren Sammlung in der Präsentation für die digitalisierten Bestände der Staats- und Universitätsbibliothek Hamburg<sup>156</sup>. Andererseits sollen die Projekte eine OAI-PMH-

<sup>155</sup> Das »MODS-DFG-Standard-Set (Bibliotheksbestände)« macht allerdings keine Vorgaben, wie die von den Digitalisierungsprojekten definierten Sammlungen in den Metadaten ausgewiesen werden sollen, vgl. Deutsche Forschungsgemeinschaft (12/2016, Anhang A).

<sup>156</sup> vgl. <http://digitalisate.sub.uni-hamburg.de/hamburgensien.html>.

Schnittstelle<sup>157</sup> zur Bereitstellung von Metadaten an übergreifende Anwendungen implementieren<sup>158</sup>. Die Sammlungsnamen werden im Kontext dieser Schnittstellen zur Bildung von Datensets genutzt<sup>159</sup>.

### Indexterme aus kontrolliertem Vokabular

Die Vokabulare, aus denen die Elementwerte stammen, sind mangels eindeutiger Referenzen oft nicht identifizierbar. Das Authority-Attribut enthält nur drei Codes, die in der Liste »Classification Scheme Source Codes« der Library of Congress verzeichnet sind<sup>160</sup>. Die unter dem Code IVDC geführte »Intranda Viewer Digital Collection Classification« ist eine von den Anwendern des Softwarepaketes Goobi der Firma Intranda GmbH nach Bedarf erstellte Hierarchiestruktur von Sammlungen<sup>161</sup>, die keinem Standard für Dokumentationssprachen entspricht. Die IVDC gehört daher zu den hauseigenen Vokabularen und zählt nicht als Quelle für kontrollierte Indexterme in dieser Datenauswertung. Die wenigen mit dem Code »sswd« ausgewiesenen Elementwerte sind Schlagwörter der GND und keine Sachgruppen der GND-Systematik, wie es die Liste der Source Codes vorgibt<sup>162</sup>. Als einziges Klassifikationssystem wird zu einem geringen Anteil die DDC verwendet.

**Tabelle 4.8: Kontrollierte Vokabulare im MODS-Element »Klassifikation«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Dewey-Dezimalklassifikation	ddc	128	3,2 %	0
Gemeinsame Normdatei	sswd	20	0,5 %	0

## 4.2.2 Themenschlagwort

**Tabelle 4.9: Datenauswertung des MODS-Elementes »Sachschlagwort«**

MODS-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/ Attribut	insgesamt	kontrolliert	unterscheidbar
<topic>	1.329	1.329	1.570	224	863

<sup>157</sup> Open Archive Initiative – Protocol for Metadata Harvesting 2.0.

<sup>158</sup> vgl. Deutsche Forschungsgemeinschaft (12/2016, S. 44).

<sup>159</sup> Diese Information stammt von Stefanie Rühle, Fachstelle Bibliothek der DDB und Ansprechpartnerin für das Zentrale Verzeichnis Digitalisierter Drucke (E-Mail am 08.08.2018).

<sup>160</sup> vgl. Library of Congress - Network Development and MARC Standards Office (13. April 2018).

<sup>161</sup> Diese Information stammt aus einem E-Mailverkehr mit Mitarbeitern der Intranda GmbH (Steffen Hankiewicz am 29.08.2017, Robert Sehr am 11.09.2017). Ein Beispiel für die Anwendung der IVDC bietet der Sammlungsbrowser des »Open Access Repository Kassel« (ORKA) unter der URL: <http://orka.bibliothek.uni-kassel.de/viewer/browse/>.

<sup>162</sup> Allerdings führt der in der Liste der Quellenangaben hinterlegte Link für die »Subject Classes of the Schlagwortnormdatei« zur Website der Gemeinsamen Normdatei und nicht zu den GND-Sachgruppen.



valueURI	127	127	224		96
authority	121	121	215	1	1
authorityURI	0		0		
Zusammenfassung	Datensätze mit mindestens einem MODS-Element »Sachschlagwort«: 45 Prozent		Indexterme aus kontrolliertem Vokabular: 14,3 Prozent, davon Indexterme mit URI: 14,3 Prozent		

### Auslegung der Definition des Metadatenelementes

Das MODS-Element »Sachschlagwort« wird selten für die Beschreibung des Themas der digitalisierten Dokumente und überwiegend nicht standardkonform verwendet. Die beiden häufigsten Elementwerte »TIFF, Vers.6.0, 600 dpi, 1 bit (s/w),ITU group; Digitalisierungsvorlage: Original« (Anzahl: 366) und »TIFF, Vers.6.0, 600 dpi, 1 bit (s/w), ITU group; Digitalisierungsvorlage Mikrofilm« (Anzahl: 75) entsprechen technischen Formatangaben. Bei kursorischer Durchsicht der weniger häufigen Elementinhalte fallen viele sogenannte Autoren-Titel-Schlüssel<sup>163</sup> auf, von denen »mathan« (Anzahl: 10) am häufigsten indexiert wurde. Dahingegen wurden nur wenige Sachschlagwörter, wie zum Beispiel »Malerei« (Anzahl: 15), von einzelnen Datengebern ausgewiesen.

### Indexterme aus kontrolliertem Vokabular

Die wenigen kontrollierten Sachschlagwörter in der MODS-Stichprobe entstammen der GND und sind mit dem URI des GND-Sachbegriffes eindeutig identifiziert. Die Quellenangabe im Authority-Attribut ist damit zwar überflüssig, wird von den Datengebern aber oft vorgenommen.

**Tabelle 4.10: Kontrollierte Vokabulare im MODS-Element »Sachschlagwort«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Gemeinsame Normdatei	gnd	224	14,3 %	224

## 4.2.3 Objektgattung

**Tabelle 4.11: Datenauswertung für das MODS-Element »Gattungsbegriff«**

MODS-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/ Attribut	insgesamt	kontrolliert	unterscheidbar

<sup>163</sup> Dabei handelt es sich um Suchschlüssel, die in den bibliothekarischen Verbundsystemen indexiert werden.

<genre>	510	510	595	415	137
valueURI	0		0		
authority	366	366	414	3	4
authorityURI	0		0		
Zusammenfassung	Datensätze mit mindestens einem MODS-Element »Gattungsbegriff«: 16,1 Prozent		Indexterme aus kontrolliertem Vokabular: 69,8 Prozent, davon Indexterme mit URI: 0 Prozent		

### Auslegung der Definition des Metadatenelementes

Das MODS-Element »Gattungsbegriff« wird, wenn es denn in den Datensätzen vorhanden ist, nicht immer für die Angabe des Dokumenttypes gebraucht. Die drei häufigsten Elementwerte »Auktionskatalog« (Anzahl: 64), »Theaterzettel« (Anzahl: 44) und »book« (Anzahl: 27) kennzeichnen den Dokumenttyp entweder nach der Art seines Inhaltes (z. B. »Auktionskatalog«) oder nach seiner Form (z. B. »book«). Bei einem signifikanten Anteil von etwa zehn Prozent der Indexterme ist jedoch entgegen den bibliothekarischen Erschließungsregeln kein Bezug zum Dokumenttyp erkennbar<sup>164</sup>. Dabei handelt es sich um Schlagwörter, die Sachgebiete ausdrücken und das Thema des Druckwerkes bezeichnen, wie zum Beispiel »Medizin« (Anzahl: 14). Dass Bibliotheken den »Gattungsbegriff« auch als Element der Inhaltserschließung interpretieren, geht auf die Verwendung der Liste der Gattungs- und Sachbegriffe der Arbeitsgemeinschaft Alter Drucke (AAD) zurück, die im Regelwerk »Alte Drucke« der Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG)<sup>165</sup> und im »MODS-Anwendungsprofil für digitalisierte Medien« empfohlen wird<sup>166</sup>.

### Indexterme aus kontrolliertem Vokabular

Die Gattungs- und Sachbegriffe der AAD werden von den MODS-Datengebern am häufigsten verwendet<sup>167</sup>. In dieser Liste sind Forms Schlagwörter (z. B. »Adressbuch«) und Sachs Schlagwörter (z. B. »Geldwesen«) nicht voneinander abgegrenzt, was im Endeffekt dazu führt, dass im MODS-Element »Gattungsbegriff« kategorial unterschiedliche Terme indexiert sind, die sich auf den Dokumenttyp

<sup>164</sup> vgl. § 5 zur Erschließung von Formangaben und Gattungsbegriffen im Regelwerk für die Schlagwortkatalogisierung (RSWK), Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (März 2017), und § 7.1.2.3 zur Erfassung der Art des Inhaltes einer Ressource im Regelwerk Resource Description and Access (RDA), Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (28. November 2017).

<sup>165</sup> vgl. Verbundzentrale des Gemeinsamen Bibliotheksverbundes (19. Juni 2013, 2.7.1 Gattungs- und Sachbegriffe).

<sup>166</sup> vgl. Rühle, Jahnke und Kühle (2017, S. 16).

<sup>167</sup> Die Liste ging 2011 aus einer Zusammenführung verschiedener Gattungsbegriffslisten im Kontext von Digitalisierungsprojekten für Alte Drucke hervor. Den Entstehungshintergrund der AAD-Gattungsbegriffsliste beschreibt Betz (2013) in seiner Masterarbeit, in deren Rahmen er den Nutzen der Vergabe von Gattungsbegriffen bei der bibliothekarischen Inhaltserschließung von Alten Drucken untersuchte.

einerseits und das Thema andererseits beziehen<sup>168</sup>. Die Indexterme der beiden anderen verwendeten Schlagwortlisten weisen den Dokumenttyp zuverlässig aus<sup>169</sup>.

**Tabelle 4.12: Kontrollierte Vokabulare im MODS-Element »Gattungsbegriff«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Gattungs- und Sachbegriffe der Arbeitsgemeinschaft Alter Drucke	aad, VD17	336	56,5 %	0
MARC Genre Term List	marcgt	54	9,1 %	0
DIN 31631-4 Kriterienkatalog für Dokumente; Codes für Einträge zu Datenkategorien	din31631-4	4	0,7 %	0

### 4.3 Encoded Archival Description (EAD)

EAD ist ein internationaler Standard, der von der Library of Congress herausgegeben wird. Das EAD-Metadatenformat in XML dient der Abbildung der Ergebnisse einer mehrstufigen archivischen Erschließung in kodierter Form. Für Datenlieferungen aus Archiven an die DDB wird das von einer Arbeitsgruppe unter Leitung des Landesarchivs Baden-Württemberg entwickelte Anwendungsprofil EAD(DDB) verwendet<sup>170</sup>.

Die Stichprobe enthält 11.018 Datensätze im EAD-Format von 89 Datengebern der Sparte »Archiv«. Mit einem Anteil von 55,1 Prozent in der Stichprobe und 57,6 Prozent in der Grundgesamtheit bilden die EAD-Datensätze die größte Menge in der DDB insgesamt.

Das EAD-Format bietet jeweils die Möglichkeit, im Source-Attribut das Vokabular anzugeben, dem der Indexterm entnommen wurde, im Normal-Attribut die kontrollierte Ansetzungsform eines Indexterms und im Authfilenumber-Attribut den Identifikator eines Normdatensatzes auszuweisen. Von diesen Attributen macht jedoch mit Ausnahme des Attributs zur Angabe der normierten Ansetzungsform des Archivalientyps kein Datengeber Gebrauch.

<sup>168</sup> Siehe Arbeitsgemeinschaft Alte Drucke beim GBV (11. Juli 2011). Die AAD-Gattungsbegriffe werden von der Verbundzentrale des GBV demnächst als RDF-Repräsentation angeboten, siehe Gemeinsamer Bibliotheksverbund (9. August 2018).

<sup>169</sup> Für die MARC Genre Term List siehe Library of Congress - Network Development and MARC Standards Office (20. Oktober 2017).

<sup>170</sup> siehe Krauth (8. Januar 2013); Landesarchiv Baden-Württemberg (2013)

### 4.3.1 Themenschlagwort

**Tabelle 4.13: Datenauswertung für das EAD-Element »Gegenstand«**

EAD-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/ Attribut	insgesamt	kontrolliert	unterscheidbar
<subject>	614	614	1.404	0	1.204
authfilenumber	0		0		
normal	0		0		
source	0		0		
Zusammenfassung	Datensätze mit mindestens einem EAD-Element »Gegenstand«: 5,6 Prozent		Indexterme aus kontrolliertem Vokabular: 0 Prozent		

#### Auslegung der Definition des Metadatenelementes

Das EAD-Element »Gegenstand« wird von den Datengebern kaum zur Beschreibung des Inhaltes einer Archivalie durch ein Sachschlagwort benutzt. Die häufigsten Werte »Patronat St. Pauli (mit Landherrenschaft Hamburger Berg)« (Anzahl: 10), »Landherrenschaft Hamburger Berg« (Anzahl: 10), und »Landherrenschaften – Hauptregistratur« (Anzahl: 8) bezeichnen Namen von Gebietskörperschaften. Andere Individualnamen, die im Element indexiert sind, sind solche von Bauwerken (z. B. »Sankt Katharinen-Kirche«, Anzahl: 3), Personen (z. B. »Vierordt, Alfred«, Anzahl: 1) und Ereignissen (z. B. »100-Jahrfeier der TH Karlsruhe«, Anzahl: 1). Auch Forms Schlagwörter sind unter den Elementwerten zu finden, zum Beispiel »Pläne« (Anzahl: 6).

### 4.3.2 Objektgattung

**Tabelle 4.14: Datenauswertung für das EAD-Element »Genre- und Formbegriffe«**

EAD-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/Attribut	insgesamt	kontrolliert	unterscheidbar
<genreform>	6.802	5.321	8.450	0	44
authfilenumber	0		0		
normal	6.590	6.590	10.628	10.628	10
source	0		0		
Zusammenfassung	Datensätze mit mindestens einem EAD-Element »Genre- und Formbegriffe«: 61,7 Prozent		Indexterme aus kontrolliertem Vokabular: 55,7 Prozent, davon Indexterme mit URI: 0 Prozent		

#### Auslegung der Definition des Metadatenelementes

Das EAD-Element »Genre- und Formbegriffe« wird von den Datengebern standardkonform für die Kennzeichnung des Archivalientyps verwendet. Die häufigsten Elementwerte sind »Schriftgut« (Anzahl: 5.096), »Akten« (Anzahl: 936) und »Spruchkammerakten« (Anzahl: 558). Unter Indextermen finden sich Bezeichnungen für generische Einfachbegriffe (z. B. »Akten«), Differenzbegriffe (z. B. »Spruchkammerakten«) und Begriffskombinationen (z. B. »Akten und Einzelstücke«). Es gibt nur wenige unterscheidbare Werte im Element und im zugehörigen Attribut für den normierten Archivalientyp.

#### Indexterme aus kontrolliertem Vokabular

Die kontrollierten Indexterme entstammen der Werteliste für das Attribut des normierten Archivalientyps, die im EAD(DDB)-Profil vorgegeben und in Erfassungssystemen der Archive hinterlegt ist<sup>171</sup>. Diese Werte fassen verschiedenartige Archivalien in Gruppen zusammen, zum Beispiel »Amtsbücher, Register und Grundbücher«. Bei der Erfassung der Archivalien mit dieser Werteliste kommt es zu Vergrößerungen und semantischen Unscharfen. Beispielsweise bezieht sich das Polysem »Pläne« in der Präkombination »Karten und Pläne« auf das kartografische Material. Die entsprechenden in der Stichprobe verzeichneten Archivalien sind aber Baupläne und Bauzeichnungen<sup>172</sup>.

<sup>171</sup> Diese Information stammt von Oliver Götz, Fachstelle Archiv der DDB (E-Mail am 08.08.2018).

<sup>172</sup> Ein Beispiel liefert das DDB-Objekt »Nord- und Südblock.- Lageplan, Vorentwurf: Bd. 2«, URL: <http://www.deutsche-digitale-bibliothek.de/item/XLVDM5TGU2ZCXRFXKAXF7BZRZKAN3X7> (letzte Aktualisierung: 20.04.2018, 15:33 Uhr).

**Tabelle 4.15: Kontrolliertes Vokabular im EAD-Element »Genre- und Formbegriffe«**

Kontrolliertes Vokabular	Anzahl der Indexterme		
	insgesamt	in Prozent	mit URI
Werteliste für den normierten Archivalientyp im EAD(DDB)-Profil 1.1	10.628	55,7 %	0

#### 4.4 Lightweight Information Describing Objects (LIDO)

Das XML-basierte Metadatenschema LIDO wurde vom International Council of Museums (ICOM) als Austauschformat für die Museumsdokumentation entwickelt. Das Format ermöglicht eine umfassende Beschreibung verschiedenartiger Objekte aus diversen Sammlungsbereichen, zum Beispiel Kunst- und Kulturgeschichte, Technik oder Naturkunde. Für Datenlieferungen an die DDB wird das von der Fachstelle Museum am Institut für Museumsforschung entwickelte Anwendungsprofil »DDB-LIDO« genutzt<sup>173</sup>. Die Datenbestände der Bildarchive und kunsthistorischen Institute orientieren sich am »Gemeinsamen Feldkatalog Graphischer Sammlungen«, der vom Deutschen Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg erstellt wurde<sup>174</sup>.

Die Stichprobe enthält 1.493 Datensätze im LIDO-Format von 62 Datengebern aus fünf Sparten: Museum (50), Forschung (4), Archiv (3), Mediathek (3) und Bibliothek (2). Der Anteil der LIDO-Datensätze beträgt 7,5 Prozent in der Stichprobe und 7 Prozent in der Grundgesamtheit aller DDB-Datensätze.

In LIDO enthalten vor allem die Elemente »Object/Work Type«, »Classification« und »Subject Concept« Indexbegriffe für die vorgeschlagenen Filter. Als Werte für einen Indexbegriff können der Identifikator (»Concept ID«) und/oder eine oder mehrere Bezeichnungen (»Term/Label«) angegeben werden. Beide Elemente, »Concept ID« und »Term/Label«, sind wiederholbar und nicht verpflichtend.

LIDO deklariert zur weiteren Differenzierung der Elemente und Elementinhalte verschiedene Attribute: (a) zur Typisierung von »Classification, »Subject Concept« und »Object/Work Type« (»type«), (b) zur Unterscheidung bevorzugter oder alternativer Bezeichnungen (»pref«), (c) zur Kennzeichnung zusätzlicher Suchterme (»addedSearchTerm«), (d) zur Typisierung von Identifikatoren (»type«), (e) zur Quellenangabe eines Begriffes (»source«) u. a. m. Diese Attribute sind im LIDO-Schema nicht immer elementspezifisch definiert. Außerdem fehlen Anwendungsregeln und kontrollierte Terme, die als Attributwerte zu verwenden sind<sup>175</sup>. In dieser Untersuchung

<sup>173</sup> siehe Fachstelle Museum der Deutschen Digitalen Bibliothek im Institut für Museumsforschung (2014).

<sup>174</sup> siehe Fichtl et al. (2015).

<sup>175</sup> Dieses Desiderat ist seit der Veröffentlichung des LIDO-Schemas im November 2010 bekannt, vgl. Stein (26. September 2011, Folie 50 ff.). Eine Arbeitsgruppe definiert seit dem Jahr 2015 kontrollierte Terme für typisierende Elemente und Attribute im LIDO-Schema Version 1.0. Die Arbeiten an den LIDO-Termen, die für diese Untersuchung relevant sind, sind noch nicht abgeschlossen, wie aus der

konnten aufgrund erheblicher Anwendungsunterschiede in den LIDO-Instanzen der verschiedenen Datengeber nicht alle der genannten Attribute berücksichtigt werden.

#### 4.4.1 Sachkategorie

**Tabelle 4.16: Datenauswertung für das LIDO-Element »Classification«**

LIDO-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/Attribut	insgesamt	kontrolliert	unterscheidbar
<classification>	1.299	1.233	2.261		
type	1.202	1.202	2.210		13
<term>	1.299	1.233	2.505	129	244
<conceptID>	164	161	245	151	110
type	164	164	250		6
source	91	91	113		6
Zusammenfassung	Datensätze mit mindestens einem LIDO-Element »Classification«: 82,6 Prozent		Indexbegriffe aus kontrolliertem Vokabular: 5,7 Prozent, davon Indexbegriffe mit URI: 5,7 Prozent		

#### Auslegung der Definition des Metadatenelementes

LIDO definiert das Element »Classification«, um Objekte aufgrund gemeinsamer Merkmale in Kategorien einer Systematik oder eines Klassifikationssystems ordnen zu können. In den DDB-Objekten enthält das Element »Term/Label« aber vor allem Bezeichnungen für Medientypen (z. B. »IMAGE«, Anzahl: 1.029), Dokumenttypen (z. B. »Druckgrafik«, Anzahl: 144) und Kunstgattungen (z. B. »Fotografie«, Anzahl: 109) sowie für Sammlungsnamen (»Fotografien der Gesellschaft S.T.E.R.N zur Altbauanierung in Kreuzberg«, Anzahl: 6) und stilistischen Epochenbezeichnungen (z. B. »Renaissance«, Anzahl: 1). Dieser Gebrauch von »Classification« in LIDO-Instanzen entspricht jedoch nicht der ursprünglich in den CDWA definierten Bedeutung<sup>176</sup>.

Auch die dreizehn Typisierungen von »Classification« lassen keine Rückschlüsse auf die Art der Klassifizierung zu. Die Werte im »Type«-Attribut überschneiden sich oft und können sich zum Beispiel auf die Art eines Begriffes (»Sachgruppe«), den Typ eines Begriffssystems (»Objektklassifikation«) oder eine spezifische Objekteigenschaft (»Language«) beziehen. Zudem

Website »Terminology« der LIDO Working Group des ICOM International Committee for Documentation hervorgeht: <http://network.icom.museum/cidoc/working-groups/lido/lido-technical/terminology/>.

<sup>176</sup> Allerdings legen unscharfe Formulierungen in den Anwendungshinweisen (»How to record«) des LIDO-Schemas und Empfehlungen in »Implementing LIDO« diese abweichende Auslegung nahe, siehe McKenna und Stein (2015, S. 7 und 39 ff.).

werden diese Typisierungen unterschiedlich von den Datengebern verwendet: Der Indexbegriff ›Keramik‹ (8) beispielsweise wird entweder als ›Sachgruppe‹ (4), als ›Gattung‹ (2) oder als ›Schlagwort‹ (2) typisiert.

### Indexterme aus kontrolliertem Vokabular

Die im LIDO-Element »Classification« indexierten Terme entstammen, soweit ersichtlich, keiner Klassifikation oder Systematik. Wenn kontrollierte Vokabulare bei der Indexierung verwendet wurden, handelt es sich meist um Schlagwörter aus Normdateien und Thesauri. Ein Teil der Vokabulare ist allerdings nicht identifizierbar, weil entweder die Quelle nicht eindeutig ausgewiesen ist<sup>177</sup> oder der URI für den Indexterm nicht aufgelöst werden kann<sup>178</sup>.

Im »Classification«-Element sind teilweise unterschiedliche Indexterme zu einem Indexbegriff zusammengefasst. Für die Sachgruppe ›Bild- und Schrifftkarten‹ zum Beispiel sind in einer Elementinstanz ein lokaler Identifikator und drei URIs ausgewiesen, die jeweils einen eigenen Begriff in der Gemeinsamen Normdatei (GND) referenzieren. Die Beziehungen der Deskriptoren untereinander und zum Indexbegriff sind nicht ausgewiesen und es ist nicht klar, ob es sich um Schnittmengen oder um Vereinigungsmengen handelt.<sup>179</sup> Dass sich lediglich der lokale Identifikator ›1.521‹ auf den Indexterm ›Bild und Schrifftkarten‹ bezieht, geht aus der Struktur der wiederholbaren Elemente »Concept ID« und »Term/Label« nicht hervor.

Zur Ermittlung der Anzahl der Indexterme aus kontrollierten Vokabularen wurden in der untenstehenden Tabelle daher die Werte im Element »Concept ID« ausgezählt. Der Prozentwert bezieht sich wiederum auf die Anzahl der Indexbegriffe im Element »Classification« (2.261).

**Tabelle 4.17: Kontrollierte Vokabulare im LIDO-Element »Classification«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Gemeinsame Normdatei	GND	148	6,5 %	148
Art & Architecture Thesaurus	AAT	2	0,1 %	2

<sup>177</sup> z. B. ist die Quellenangabe »xTree« der Name eines webbasierten Vokabularverwaltungswerkzeuges, das vom digiCULT-Verbund eG entwickelt wurde, siehe <http://www.digicult-verbund.de/de/digicultxtree>. Dahinter verbergen sich in der Regel Systematiken, wie etwa die Hessische Systematik.

<sup>178</sup> Dies betrifft v. a. URIs mit dem Namensraum »<http://digicult.vocnet.org/>«, z. B. »<http://digicult.vocnet.org/stil/stil00071>«.

<sup>179</sup> Der im LIDO-Element »Concept ID« indexierte URI »<http://d-nb.info/gnd/4744722-9>« verweist in der GND auf den Indexterm ›Bildkarte‹, »<http://d-nb.info/gnd/4419132-7>« auf den Indexterm ›Schriftgut‹ und »<http://d-nb.info/gnd/4444708-5>« auf den Indexterm ›handgezeichnete Karte‹.



## 4.4.2 Themenschlagwort

**Tabelle 4.18: Datenauswertung für das LIDO-Element »Subject Concept«**

LIDO-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/Attribut	insgesamt	kontrolliert	unterscheidbar
<subjectConcept>	360 (531)	360 (531)	909 (1.521)		
type	788 (531)	788 (531)	593 (1.521)		3 (1)
<term>	360 (531)	360 (531)	1.259 (1.521)	610 (0)	1.014 (520)
<conceptID>	271 (0)	271	787 (0)	785	576
type	271 (0)	271	787 (0)		2
source	271 (0)	271	753 (0)		8
Zusammenfassung	Datensätze mit mindestens einem LIDO-Element »Subject Concept«: 24,1 Prozent		Indexbegriffe aus kontrolliertem Vokabular: 67,1 Prozent, davon Indexbegriffe mit URI: 58,0 Prozent		

### Auslegung der Definition des Metadatenelementes

Das LIDO-Element »Subject Object« für Individualbegriffe wird von einigen Datengebern abweichend von der LIDO-Definition verstanden und mit Werten für Allgemeinbegriffe belegt, die definitionsgemäß in »Subject Concept« gehören<sup>180</sup>. In der Auswertungstabelle oben stehen die aus den entsprechenden Instanzdaten ermittelten Zahlen<sup>181</sup> daher in Klammern hinter den Zahlen für die Elementwerte von »Subject Concept«.

Die in »Subject Concept« enthaltenen Werte beschreiben in der Regel das Thema der Museumsobjekte. Am häufigsten sind die Werte ›Weltkrieg, 1.« (25), ›Weltkrieg, 2.« (25) und ›Nationalsozialismus« (22) indexiert. Die Mehrzahl der Indexbegriffe ist entweder als ›Schlagwort« (1.521), ›Beschreibung« (333), ›Ikonographie« (254) oder ›Interpretation« (6) typisiert. Die mit ›Ikonographie«, ›Beschreibung« und ›Interpretation« typisierten Begriffe beziehen sich auf das Thema von Werken der bildenden Kunst; ›Ikonographie« meint die Indexierung mit Iconclass durch eine Notation, zum Beispiel ›11 F 41 21« mit der zugehörigen Bezeichnung ›Maria stehend (oder halbfigurig), das Christuskind steht auf ihrem Arm«. Das Element »Subject Concept« enthält auch Schlagwörter, die das Thema eines Objektes spezifisch benennen (z. B. ›Weltkrieg, 1.«, 25), die allgemein den Sachbereich beschreiben (z. B. ›Landwirtschaft«, 8) oder biologische Gattungen

<sup>180</sup> Zum Beispiel kann »Subject Object« ein bestimmtes Bauwerk, das auf einem Foto abgebildet ist, benennen oder bevorzugt durch einen URI auf eine Beschreibung des entsprechenden Bauwerkes in einer Normdatei referenzieren.

<sup>181</sup> Hierbei handelt es sich um die Auswertung der LIDO-Daten der Stiftung Deutsches Historisches Museum.

referenzieren (z. B. ›Vögel‹, 1). Einige Schlagwörter weisen die Art des Werkes anhand seines Inhaltes aus, zum Beispiel ›Bildnis‹ (17) und ›Männerportrait‹ (3).

### Indexterme aus kontrolliertem Vokabular

Der überwiegende Teil der Indexterme stammt aus verbreiteten Vokabularen. Am häufigsten verwendet wird das in der Museumsdokumentation bekannte Klassifikationssystem Iconclass<sup>182</sup>. Auch die bibliothekarischen Normdateien Gemeinsame Normdatei (GND) und Library of Congress Subject Headings (LCSH) werden genutzt. Auffallend wenig Anwendung findet der in den LIDO-Anwendungsprofilen empfohlene Art & Architecture Thesaurus (AAT). Weitere museumsspezifische Vokabulare, die zur Verschlagwortung verwendet werden, sind die Grobssystematik des Instituts für Museumskunde (1988) und die Oberbegriffsdatei der Landesstelle für die nichtstaatlichen Museen in Bayern (2011)<sup>183</sup>. Auch im Element »Subject Concept« finden sich Indexterme mit nicht auflösbaren URIs<sup>184</sup> und ohne identifizierbare Quellen, die in dieser Datenauswertung nicht zu den kontrollierten Vokabularen gezählt werden können.

Auch wenn alle verwendeten Vokabulare, außer der Grobssystematik, als Linked Open Data publiziert sind, weisen die Datengeber nicht immer den URI des Indexbegriffes in den LIDO-Instanzen aus. Vielfach ist im Element »Concept ID« ein Identifikator eines Normdatensatzes angegeben und im Attribut »Source« eine nicht normierte Quellenangabe hinterlegt<sup>185</sup>. Derartige Identifikatoren wurden in dieser Datenauswertung nicht als URI gezählt.

**Tabelle 4.19: Kontrollierte Vokabulare im LIDO-Element »Subject Concept«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Iconclass	<a href="http://iconclass.org">http://iconclass.org</a> , Iconclass	518	57 %	504
Gemeinsame Normdatei	GND	135	14,9 %	32
Library of Congress Subject Headings	LCSH	64	7 %	0
Art & Architecture Thesaurus	aat	20	2 %	0
Grobssystematik	Grobssystematik	19	2 %	0
Oberbegriffsdatei	oberbegriffsdatei	15	1,7 %	0

<sup>182</sup> siehe Rijksbureau voor Kunsthistorische Documentatie (2012).

<sup>183</sup> Eine Beschreibung dieser Vokabulare bietet die Website <http://museum.zib.de/museumsvokabular/index.php?main=download&ls=9&co=we> der Initiative museumsvokabular.de (2007).

<sup>184</sup> z. B. ›<http://digicult.vocnet.org/ikonographie/3.1058>‹

<sup>185</sup> z. B. ist in einer Elementinstanz von »Subject Concept« in der Stichprobe folgende Kombination zu finden: Identifikator ›4040457-2‹ mit der Quellenangabe ›gnd‹, ›sh85085422‹ mit der Quellenangabe ›lcsch‹, ›8027‹ mit der Quellenangabe ›Grobssystematik‹, ›7125‹ mit der Quellenangabe ›oberbegriffsdatei‹ und ›300264626‹ mit der Quellenangabe ›aat‹.

### 4.4.3 Objektgattung

**Tabelle 4.20: Datenauswertung für das LIDO-Element »Object/Work Type«**

LIDO-Element und Attribut	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element/Attribut	mit Werten im Element/Attribut	insgesamt	kontrolliert	unterscheidbar
<objectWorkType>	1.493		1.899		
type	1.020	1.020	1.296		5
<term>	1.493	1.493	2.637	546	619
<conceptID>	372	372	555	555	104
type	372	372	555		1
source	372	372	555		7
Zusammenfassung	Datensätze mit mindestens einem LIDO-Element »Object/Work Type«: 100 Prozent		Indexbegriffe aus kontrolliertem Vokabular: 28,8 Prozent, davon Indexbegriffe mit URI: 28,8 Prozent		

#### Auslegung der Definition des Metadatenelementes

Das Element »Object/Work Type« wird als Pflichtelement in LIDO von den Datengebern meistens standardkonform verwendet. »Foto« (114), »graphic arts | Grafik« (83) und »Druck« (65) sind die häufigsten Werte in diesem Element. Diese Indexterme kennzeichnen den Typ des Objektes unter verschiedenen Aspekten: zum Beispiel nach Material und Herstellungstechnik (z. B. »Foto«) oder nach Form und Funktion (z. B. »graphic arts | Grafik«). Einige Datengeber drücken anhand der Werte im Typ-Attribut verschiedene Ebenen der Indexierungsspezifität aus: Zum Beispiel ist »Ansichtspostkarte« (15) als »Sachbegriff« und »Postkarte« (23) als »allgemeiner Sachbegriff« typisiert.

#### Indexterme aus kontrolliertem Vokabular

Entgegen den Empfehlungen im LIDO-Schema stammen die meisten Indexterme für den »Object/Work Type« nicht aus einem kontrollierten Vokabular. Wenn kontrollierte Indexterme verwendet werden, sind es hauptsächlich Deskriptoren für den Objekttyp aus Spezialthesauri. In wenigen Fällen wird die bibliothekarische Normdatei GND oder die museale Oberbegriffsdatei (OBG) zur Indexierung verwendet.

In vielen Instanzen von »Object/Work Type« sind jeweils mehrere URIs für verschiedene Begriffe ausgewiesen, deren semantische Relationen zueinander in einem kontrollierten Vokabular abgebildet sind. Beispielsweise ist in der OBG der Begriff »Fotografie« durch ein Mapping mit

exakter Übereinstimmung auf die Begriffe ›photographs‹ im AAT und ›Fotografie‹ in der GND abgebildet<sup>186</sup>. In den entsprechenden LIDO-Instanzen sind diese Begriffe jedoch bezugslos im Element »Concept ID« referenziert<sup>187</sup>:

**Tabelle 4.21: Kontrollierte Vokabulare im LIDO-Element »Object/Work Type«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Partage <sup>188</sup>	<a href="http://partage.vocnet.org/PPObjektType">http://partage.vocnet.org/PPObjektType</a> , <a href="http://partage.vocnet.org/PPRefiningObjektType">http://partage.vocnet.org/PPRefiningObjektType</a>	286	15,3 %	286
Art & Architecture Thesaurus	AAT, <a href="http://vocab.getty.edu/aat/">http://vocab.getty.edu/aat/</a>	228	12,2 %	228
Gemeinsame Normdatei	GND	37	2 %	37
Oberbegriffsdatei	xTree	4	0,2 %	4

## 4.5 Dublin Core Metadata Element Set (DCMES)

Das von der Dublin Core Metadata Initiative (DCMI) entwickelte und gepflegte Elementset des Dublin-Core-Standards wird spartenübergreifend zur Beschreibung von Objekten unterschiedlicher Art verwendet. Das DCMES bildet die Grundlage für verschiedene Anwendungsprofile, in denen die Datengeber der DDB ihre Metadaten übermitteln. Häufig verwendete Profile sind das für die Datenlieferung an das Europeana-Portal entwickelte Europeana Semantic Element Set (ESE)<sup>189</sup> und das OAI\_DC<sup>190</sup>, ein Harvesting-Format, das zum Beispiel im Kontext von wissenschaftlichen Repositorien für elektronische Publikationen verwendet wird<sup>191</sup>.

Die Stichprobe enthält 1.217 Datensätze im Dublin-Core-Format von acht Datengebern aus vier Sparten: Bibliothek (3), Forschung (2), Mediathek (2) und Archiv (1). Der Anteil der Dublin-Core-Datensätze beträgt 6,1 Prozent in der Stichprobe und 5 Prozent in der Grundgesamtheit.

<sup>186</sup> Diese Aussage kann z. B. durch den Aufruf des OBG-URI ›<http://obg.vocnet.org/x001212x>‹ in einem Webbrowser nachvollzogen werden. Der URI dereferenziert eine menschenlesbare Präsentation des OBG-Begriffes ›Fotografie‹ im Portal »xTree.Public« (letzte Änderung: 05.12.2016).

<sup>187</sup> Für das vorgestellte Beispiel gibt es in der Stichprobe vier Datensätze, in denen die URIs ›<http://obg.vocnet.org/x001212x>‹, ›<http://vocab.getty.edu/aat/300046300>‹ und ›<http://d-nb.info/gnd/4045895-7>‹ als Werte im Element »Concept ID« in einer Instanz von »Object/Work Type« ausgewiesen sind.

<sup>188</sup> Dabei handelt es sich um einen Thesaurus, der im Rahmen des EU-Projektes »Partage Plus« für die Indexierung von Objekten des Jugendstils in der Europeana auf Basis des AAT entwickelt wurde, siehe Deutsches Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg ([o. J.]) sowie Pagel und Stein (2013).

<sup>189</sup> siehe Clayphan (14. Juli 2013).

<sup>190</sup> Im OAI-PMH-Protokoll zum Austausch von Metadaten wird das DCMES als Metadatenformat mit dem Präfix »oai\_dc« verwendet, siehe <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

<sup>191</sup> Anwendungsempfehlungen für das Format OAI\_DC bietet zum Beispiel das »DINI-Zertifikat für Open-Access-Repositorien und – Publikationsdienste 2016«, siehe DINI-Arbeitsgruppe "Elektronisches Publizieren" (S. 48 ff.).

Das DCMES bietet in der Syntax von XML keine Möglichkeit, die Quelle und den Identifikator oder URI eines Indexterms in einem Attribut explizit auszuweisen. Einige Datengeber definieren eigene Attribute für diesen Zweck; diese sind aber anwenderspezifisch und konnten im Rahmen dieser Datenauswertung nicht berücksichtigt werden.

#### 4.5.1 Sachkategorie und Themenschlagwort

**Tabelle 4.22: Datenauswertung für das DCMES-Element »Thema«**

DCMES-Element	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element	mit Werten im Element	insgesamt	kontrolliert	unterscheidbar
<subject>	1.211	1.211	4.241	232	640
Zusammenfassung	Datensätze mit mindestens einem DCMES-Element »Thema«: 95,3 Prozent		Indexterme aus kontrolliertem Vokabular: 3,1 Prozent, davon Indexterme mit URI: 0 Prozent		

##### Auslegung der Definition des Metadatenelementes

Das DCMES-Element »Thema« wird von den Datengebern nicht immer standardkonform verwendet, was schon anhand der häufigsten Elementwerte »Fotos« (1.026), »Foto« (423), »Fotografie« (409) ersichtlich ist. Insgesamt beziehen sich etwa die Hälfte der Indexterme nicht auf das Thema, sondern auf den Typ des Objektes. Weitere Beispiele sind »Kontaktbogen« (339) und »Technik-Katalog« (79). Themen sind im DCMES-Element durch Sachschlagwörter (z. B. »Architektur«, 114), Bezeichnungen für Sachgruppen oder Kategorien (z. B. »Kunst und Kultur«, 52) oder Notationen (z. B. »IconClass: 49 C 16 4«, 2) ausgedrückt.

##### Indexterme aus kontrolliertem Vokabular

Kontrollierte Vokabulare sind in den DCMES-Daten nur anhand der Quellenangabe im Elementinhalt erkennbar (z. B. »ddc:400«). Die wenigen kontrollierten Indexterme für das »Thema« sind Notationen aus den Klassifikationssystemen Iconclass und DDC.

**Tabelle 4.23: Kontrolliertes Vokabular für das DCMES-Element »Thema«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Iconclass	IconClass:	130	3,1 %	0
Dewey Decimal Classification	ddc:	2	0,1 %	0

## 4.5.2 Objektgattung

**Tabelle 4.24: Datenauswertung für das DCMES-Element »Typ«**

DCMES-Element	Anzahl der Datensätze		Anzahl der Werte		
	mit dem Element	mit Werten im Element	insgesamt	kontrolliert	unterscheidbar
<type>	1.201	1.201	1.212	6	20
Zusammenfassung	Datensätze mit mindestens einem DCMES-Element »Typ«: 94,5 Prozent		Indexterme aus kontrolliertem Vokabular: 0,5 Prozent, davon Indexterme mit URI: 0 Prozent		

### Auslegung der Definition des Metadatenelementes

Das DCMES-Element »Typ« wird einheitlich für die Angabe der Art des Objektes verwendet. Die meisten Bezeichnungen weisen einen allgemeinen Typ aus, wie die beiden häufigsten Werte ›image‹ (1.175) und ›sound‹ (12). Spezifischere Typen wie ›Grafik‹ (5) oder ›Zeichnung‹ (2) sind kaum vorhanden.

### Indexterme aus kontrolliertem Vokabular

Anhand der Quellenangabe im Elementinhalt ist nur ein Vokabular erkennbar, bei dem es sich um eine Schlagwortliste für Dokumenttypen aus dem Bereich des elektronischen Publizierens handelt (z. B.: ›doc-type:bookpart‹)<sup>192</sup>.

**Tabelle 4.25: Kontrolliertes Vokabular im DCMES-Element »Typ«**

Kontrolliertes Vokabular	Quellenangabe	Anzahl der Indexterme		
		insgesamt	in Prozent	mit URI
Gemeinsames Vokabular für Publikations- und Dokumenttypen der DINI-AG „Elektronisches Publizieren“	doc-type:	6	0,5 %	0

<sup>192</sup> vgl. DINI-Arbeitsgruppe "Elektronisches Publizieren" (2010).

## 5 Fazit und Ausblick

Diese Arbeit ist von der Beobachtung ausgegangen, dass die Retrievalmöglichkeiten in der Deutschen Digitalen Bibliothek (DDB) nicht befriedigend sind. Es fehlen vor allem navigierbare Sucheinstiege und wirksame Filtermöglichkeiten. Besonders für informationsorientierte Suchen und zur explorativen Erkundung sollte die DDB mehr Unterstützung anbieten. Deshalb wurden in Kapitel 2.3 Vorschläge für drei Informationsfilter entwickelt, die spartenübergreifend fundamentale Facetten abdecken und ein nutzerfreundliches Browsen im DDB-Portal ermöglichen könnten. Die erfolgreiche Anwendung dieser Filter setzt jedoch voraus, dass bestimmte Metadatenelemente in den gelieferten Objekten vorhanden und mit geeigneten Werten belegt sind. Im empirischen Teil dieser Arbeit (Kapitel 4) wurde deshalb untersucht, ob und wenn ja inwieweit die Metadaten der gelieferten Objekte in der DDB die Voraussetzungen für ein filterbasiertes Retrieval mitbringen. Die Untersuchung ging von zwei Fragen an die Metadaten der DDB-Objekte aus:

1. Definieren die einzelnen Metadatschemata übereinstimmend Elemente, die sich spartenübergreifend als Filter eignen (siehe Kapitel 3.2)?
2. Eignen sich die Inhalte der Metadatenelemente als Navigationselemente (siehe Kapitel 4)?

Zu 1.: Das Erstellen von Konkordanzen zwischen den spartenspezifischen Schemata zeigte, dass die Elemente für Objekttyp, inhaltliche Indexierung und systematische Kategorisierung semantisch weitgehend korrespondieren. Sie konnten nahtlos auf die vorgeschlagenen drei Filter »Objektgattung«, »Themenschlagwort« und »Sachkategorie« abgebildet werden.

Zu 2.: Die Elementinhalte wurden quantitativ und qualitativ auf ihre Eignung als Selektionsmerkmale bzw. Filterwerte untersucht. Als Korpus diente eine randomisierte Stichprobe aus 20.000 Datensätzen über die Grundgesamtheit von etwa 20 Millionen Datensätzen der DDB zum Untersuchungszeitpunkt.

In Kapitel 5.1 werden die Ergebnisse der Schema- und Stichprobenanalysen formatübergreifend betrachtet und ein zusammenfassender Befund für jedes Untersuchungskriterium wird formuliert. Der Untersuchungsbefund führt in Kapitel 5.2 zu Überlegungen, welche Maßnahmen zur Verbesserung der Metadaten eingeleitet werden müssten, damit die vorgeschlagenen Informationsfilter in der DDB sinnvoll eingesetzt werden können. Abschließend werden Desiderate benannt, die weitere Forschungen anregen könnten.

### 5.1 Formatübergreifende Zusammenfassung der Ergebnisse

Die Untersuchung der Stichprobe ergab, dass die objektbeschreibenden Metadaten die genannten Voraussetzungen (Kapitel 2.3 und 2.4) für ein filterbasiertes Retrieval in einigen Aspekten nicht erfüllen.

Die erste und grundlegende Voraussetzung, wonach die Metadatschemata den vorgeschlagenen Filtern entsprechende Elemente definieren, ist weitgehend erfüllt (siehe Kapitel 3.2).

Die zweite Voraussetzung, wonach die den Filtern entsprechenden Metadatenelemente in den DDB-Instanzen vorhanden sind, ist in vielen Fällen nicht erfüllt. Da die Schemata allerdings bis auf eine Ausnahme<sup>193</sup> kein einziges Element als verpflichtend deklarieren, muss ohnehin mit einer spärlichen Verwendung der Datenelemente gerechnet werden. Die Ergebnisse der Stichprobenanalyse bestätigen diese Erwartung:

- Insgesamt enthält kaum mehr als ein Zehntel der Datensätze ein Element mit einem Wert für alle drei vorgeschlagenen Filter.
- Mehr als ein Viertel der Datensätze stellt überhaupt keine Elementwerte für einen der Filter bereit.

Für welchen der möglichen Filter die Datengeber vorwiegend Erschließungsdaten in den Metadatenformaten bereitstellen, unterscheidet sich je nach Sparte:

- Werte für die »Objektgattung« liefern alle Museen, da der Objekttyp ein Pflichtelement ist. Archive folgen mit fast einem Drittel ausgewiesener Archivalientypen für die »Objektgattung«, bei Bibliotheken liegt die Anzahl der indexierten Dokumenttypen für die »Objektgattung« unter einem Fünftel.
- 30 bis 40 Prozent der bibliothekarischen Datensätze enthalten Werte für das »Themenschlagwort«. Bei Museumsobjekten liegt dieser Anteil bei einem Viertel, bei archivischen Objekten um fünf Prozent.
- Für die »Sachkategorie« enthalten 80 Prozent der Museumsdatensätze Werte sowie zwei Drittel der bibliothekarischen Datensätze. Archive können keine Werte liefern, da das entsprechende Metadatenelement fehlt.

Die dritte Voraussetzung, wonach die Metadatenelemente mit geeigneten Deskriptoren aus kontrollierten Vokabularen belegt sein sollen, ist weitgehend nicht erfüllt<sup>194</sup>. Die qualitative Analyse zeigte, dass viele Indexterme als Werte für Suchfilter, vor allem in Aggregationsportalen, nicht geeignet sind:

- Weniger als die Hälfte aller Indexterme der Stichprobe weist ein kontrolliertes Vokabular als Quelle aus.
- Weniger als ein Zehntel der Indexterme sind durch einen URI eindeutig mit dem Begriff eines Vokabulars verbunden.
- Nur sieben der 18 ausgewiesenen Vokabulare sind als Linked Open Data (LOD) publiziert.
- Die überwiegende Zahl der kontrollierten Indexterme (93 Prozent) stammt aus Vokabularen<sup>195</sup>, die nicht spartenübergreifend verwendet werden und die auch wegen ihres engen Geltungsbereichs in Aggregationsportalen nicht vermittelnd wirken können. Die

---

<sup>193</sup> Das einzige Pflichtelement dieser Untersuchung ist »Object/Work Type« im LIDO-Schema.

<sup>194</sup> vgl. Kapitel 2.4.2 dieser Arbeit.

<sup>195</sup> siehe Anhang B: Übersicht der in der DDB verwendeten kontrollierten Vokabulare.



Hälfte der Vokabulare ist außerdem wegen struktureller Schwächen für Suchausweitungen und thematische Navigation ungeeignet.

In den bibliothekarischen Formaten sind mit insgesamt elf Prozent die Dewey-Dezimalklassifikation (DDC) und die DDC-Sachgruppen in Gebrauch. Der Art & Architecture Thesaurus (AAT) wird mit einem Anteil von nur 0,5 Prozent nicht einmal in der Museumsdokumentation häufig verwendet.

Das einzige Vokabular, das spartenübergreifend von Bibliotheken, Bildarchiven und Museen für die Erschließung benutzt wird, ist die Gemeinsame Normdatei (GND)<sup>196</sup>. Der Anteil der GND-Deskriptoren macht allerdings nur sieben Prozent in der untersuchten Stichprobe aus. Nichtsdestotrotz sollte die GND als das semantische Bindeglied für den Datenbestand der DDB in Betracht gezogen werden. Auch das Projekt »GND für Kulturdaten« (GND4C)<sup>197</sup> legt die Verwendung der GND als Daten-Drehscheibe nahe: Sie soll zukünftig ein »fächer- und spartenübergreifendes Normdaten-Hub«<sup>198</sup> sein und als »Rückgrat eines globalen maschinenlesbaren, semantischen Netzes der Kultur und Wissenschaft«<sup>199</sup> fungieren. Für die DDB würden sich daraus unmittelbare Vorteile ergeben:

Die einzelnen Sammlungen gewinnen dabei an Sichtbarkeit und sind gut für den Einsatz in spartenübergreifenden Nutzungskontexten vorbereitet. Davon profitieren insbesondere integrierende Rechercheportale wie die Deutsche Digitale Bibliothek (DDB). (Niggemann et al., 2017, S. 2f.)

## 5.2 Überlegungen für das weitere Vorgehen zur Verbesserung der Retrievalmöglichkeiten in der DDB

Die Untersuchungsergebnisse lassen vermuten, dass beim Stand der Metadatenqualität der DDB-Objekte eine filterbasierte Suche hohe Informationsverluste mit sich bringen würde. Es ist deshalb zu überlegen, wie die Datenqualität verbessert werden kann. Angesichts der Datenmengen wäre zu untersuchen, inwieweit automatische Verfahren zur Unterstützung eingesetzt werden können.

### 5.2.1 Empfehlungen an die Datengeber

In einem ersten Schritt sind ergänzende Anforderungen an die Metadatenschemata und die Inhalte der Metadatenelemente zu formulieren und den Datengebern zu übermitteln<sup>200</sup>:

1. Alle Metadatenschemata müssen die erforderlichen Elementdeklarationen enthalten.

---

<sup>196</sup> Die GND wurde in Kapitel 2.4.4 dieser Arbeit vorgestellt.

<sup>197</sup> Laufzeit: I. Phase: Mai 2018 bis April 2020. Informationen zum Projekt bietet die Wikiseite »GND für Kulturdaten (GND4C)« der Deutschen Nationalbibliothek. URL: <https://wiki.dnb.de/x/dlf9Bw> (Erstellt von Hartmann, Sarah, zuletzt geändert von Fischer, Barbara am 2018-05-31).

<sup>198</sup> vgl. GND4C-Wiki, URL: <https://wiki.dnb.de/x/dlf9Bw> (Erstellt von Hartmann, Sarah, zuletzt geändert von Fischer, Barbara am 2018-05-31).

<sup>199</sup> vgl. GND4C-Wiki, URL: <https://wiki.dnb.de/x/dlf9Bw> (Erstellt von Hartmann, Sarah, zuletzt geändert von Fischer, Barbara am 2018-05-31).

<sup>200</sup> Die Anforderungen an die Metadatenschemata müssten in Gremien oder Arbeitsgruppen eingebracht werden, die für die Weiterentwicklung der Schemata bzw. Anwendungsprofile verantwortlich sind.

2. Alle Metadatenschemata sollten die eindeutige Referenzierung von Begriffen aus kontrollierten Vokabularen mittels URIs erlauben. Die DDB sollte ihren Datengebern dringend empfehlen, die Datenelemente für die vorgeschlagenen Filter mit geeigneten Werten, bevorzugt aus spartenübergreifend verwendbaren LOD-Vokabularen, zu belegen.
3. Die Datengeber sollten darauf hingewiesen werden, dass die Metadatenelemente spezifikationskonform angewendet werden müssen, sodass vor allem die korrekte Unterscheidung zwischen dem Thema und der Gattung eines Objektes gewährleistet ist.
4. Schließlich sollte die DDB ihren Datengebern deutlich machen, dass die ausschließliche Indexierung mit freien Schlagwörtern oder hauseigenen Deskriptoren nicht nur die Retrievalqualität in Portalen erheblich mindert, sondern auch die Sichtbarkeit der eigenen Bestände herabsetzt.

## 5.2.2 Entwicklung eines Rankingsystems

In einem zweiten Schritt sollte ein Rankingsystem mit Qualitätsanforderungen entwickelt werden, das die Eignung der Lieferdaten für das Retrieval in der DDB bewertet<sup>201</sup>, angelehnt an das »Publishing Framework« der Europeana. Unter dem Motto »The more you give, the more you get« werden die Datengeber zur Optimierung ihrer Daten für Suche und Retrieval aufgefordert<sup>202</sup>. Da die DDB Datenzulieferer für Europeana ist, liegt es nahe, diese Kriterien auch als Anforderungen an die Metadaten der DDB-Objekte zu formulieren<sup>203</sup>.

Die DDB sollte darüber hinaus bei ihren Mindestanforderungen die bereits angesprochenen Unterschiede der spartentypischen Erfassungsschemata und Erschließungstraditionen berücksichtigen (siehe auch Kapitel 5.1, Seite 69):

- Alle DDB-Objekte aus der Sparte Museum sollten einen Wert für »Objektgattung« enthalten. Im Museumsbereich ist der Objekttyp<sup>204</sup> die wichtigste Information. Dagegen ist eine Inhaltsbeschreibung oft nicht möglich, weil dreidimensionale Objekte, wie Alltagsgegenstände oder naturkundliche Exemplare, gewöhnlich kein Thema haben.
- Alle DDB-Objekte aus der Sparte Bibliothek sollten einen Wert für »Objektgattung« und ein »Themenschlagwort« enthalten. In Bibliotheken steht traditionell das »Themenschlagwort« im Vordergrund. Aber auch für den Objekttyp gibt es ein verbindliches Vokabular mit ausreichender Diskriminanz<sup>205</sup>.

<sup>201</sup> Ein solches Rankingsystem könnte darüber hinaus auch als Anreiz für die Datengeber fungieren, die Qualität ihrer Daten zu verbessern, um das Retrieval und die Sichtbarkeit ihrer Objekte zu verbessern.

<sup>202</sup> vgl. Europeana Foundation (2015). Je nachdem, welche der Anforderungen die Daten in diesem dreistufigen »Tier«-System erfüllen, qualifizieren sie sich für bestimmte Anwendungsszenarien.

<sup>203</sup> Relevant sind hier die Anforderungen für das in Kapitel 2.2.4 erwähnte Retrievalszenario »Search and browse by subjects« der Europeana, siehe Charles, Isaac und Hill (2015, S. 23-25).

<sup>204</sup> Zu den spartenspezifischen Bezeichnungen für den Typ (Objekttyp, Dokumenttyp, Archivalientyp) siehe Kapitel 3.2.3 Filter »Objektgattung«.

<sup>205</sup> Die RDA-Arbeitshilfe AH-007 enthält 164 normierte Begriffe zur Beschreibung des Dokumenttyps (ehemals Forms Schlagwort) nach Art des Inhalts, siehe Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (28. November 2017).

- Alle DDB-Objekte aus der Sparte Archiv sollten einen Wert für »Objektgattung« enthalten. In Archiven hat sich eine verbale thematische Erschließung noch nicht etabliert<sup>206</sup>; für den Archivalientyp wurde jedoch ein verbindliches Vokabular erarbeitet, das auch von einem Drittel der Archive genutzt wird<sup>207</sup>.
- Alle DDB-Objekte sollten einen Wert für »Sachkategorie« enthalten. Eine systematische oder klassifikatorische Einordnung der Objekte ist unabhängig von der Sparte möglich und wird in Bibliotheken und Museen auch bereits genutzt<sup>208</sup>.

Der »Sachkategorie« kommt unter den vorgeschlagenen Filtern eine besondere Bedeutung zu:

- Sie kann spartenübergreifend gefordert werden. Das macht die »Sachkategorie« zu einem Bindeglied zwischen den spartenspezifischen Erschließungen.
- Sie ist für eine Top-down-Navigation gut geeignet. Das macht die »Sachkategorie« zu einem nutzerfreundlichen Ausgangspunkt für das explorative Suchen und Entdecken (siehe Kapitel 2.2).

In der DDB werden zur Zeit allerdings für die »Sachkategorie« eine Fülle unterschiedlicher Vokabulare verwendet, sofern überhaupt Werte geliefert werden<sup>209</sup> (siehe Kapitel 4). Um spartenübergreifend einen systematischen Zugang zum DDB-Bestand anbieten zu können, ist die gemeinsame Nutzung eines für alle Beteiligten verbindlichen Vokabulars eine notwendige Voraussetzung.

### 5.2.3 Entwicklung eines Kategoriensystems

In einem dritten Schritt ist daher ein Vokabular für die Sachkategorien der DDB zu entwickeln und bereitzustellen, das über wenige hierarchische Stufen einen Einstieg in die Suche ermöglicht<sup>210</sup>, einen ersten Überblick über das Datenangebot gibt, sich zum Browsen und Entdecken eignet und nicht zuletzt auch zum Filtern von Ergebnismengen verwendet werden kann.

Das für die DDB zu entwickelnde Kategoriensystem sollte:

1. eine hierarchische Struktur für eine Top-down-Navigation bieten,
2. Konkordanzen zu den häufig verwendeten, etablierten Klassifikationen ausweisen, vor allem zur DDC, mit der ein Großteil der bibliografischen Objekte in der DDB erschlossen ist,
3. dereferenzierbare URIs zur Anreicherung des Suchindex bereitstellen,

<sup>206</sup> Nach dem Regelwerk »Describing Archives: A Content Standard« gehört die verbale Inhaltserschließung von archivischem Material mit Themenschlagwörtern zu den wichtigsten Aspekten, um einen Zugangspunkt für die Recherche zu bieten, vgl. Society of American Archivists (2013, XXIII, »Topical Subjects«).

<sup>207</sup> Aus der Wertliste für den normierten Archivalientyp im EAD(DDB)-Profil stammen mehr als die Hälfte dieser Indexterme, siehe Kapitel 4.3.2 dieser Arbeit.

<sup>208</sup> Circa zwei Drittel der untersuchten Datensätze aus diesen Sparten enthalten einen Elementwert für »Sachkategorie«: das Bibliotheksformat MARC 68 Prozent, das Bibliotheksformat MODS 67 Prozent und das Museumsformat LIDO 83 Prozent.

<sup>209</sup> Die von den Datengebern verwendeten Klassifikationen und Systematiken erfüllen zudem nicht die Anforderungen an kontrollierte Vokabulare im Kontext der DDB: Die DDC-Notationen sind zum Beispiel nicht durch URIs dereferenzierbar und das LOD-Vokabular Iconclass deckt nur den Bedarf der ikonografischen Bilderschließung ab.

<sup>210</sup> Dass sich überschaubare Kategoriensysteme für den Sucheinstieg in umfangreichen Datenangeboten gut eignen, hat die Voruntersuchung von Informationsfiltern in Portalen des elektronischen Handels in Kapitel 2.2.3 dieser Arbeit aufgezeigt.

#### 4. Fachgebiete und Themen spartenübergreifend umfassend abdecken.

Unter den Vokabularen, die als Grundlage für ein Kategoriensystem der DDB in Frage kommen, wären unter anderen zu nennen:

- die DDC-Sachgruppen mit 101 Einträgen (siehe auch Kapitel 2.4.3),
- die GND-Sachgruppen mit 36 Kategorien auf der obersten Ebene (siehe auch Kapitel 2.2.4),
- das Wikipedia-Kategoriensystem, das sehr umfangreich, aber universell und praxisnah ist und den Erwartungen der Nutzer stärker entgegenkommt als die bibliotheksorientierte DDC,
- das Europeana Vocabulary mit 360 Termen für eine Objektfacette und sechs Sachkategorien, untergliedert in drei Hierarchiestufen<sup>211</sup>.

Diese Vokabulare könnten als Ausgangspunkte für die Entwicklung des DDB-spezifischen Kategoriensystems (im Folgenden kurz »Kategoriensystem«) herangezogen werden. Dabei sollten die einzelnen Klassen, Sachgruppen oder Kategorien der Ausgangsvokabulare auf das »Kategoriensystem« abgebildet werden.

Da die Entwicklung eines Kategoriensystems aufwendig ist, sollten Möglichkeiten einer maschinellen Unterstützung evaluiert werden. Dafür gibt das Projekt »Europeana Food and Drink« mit seiner Klassifikation, die teilautomatisch aus den Wikipedia-Kategorien gewonnen wurde, eine Fülle von Anregungen<sup>212</sup>. Aber nicht nur für die Entwicklung des »Kategoriensystems«, sondern auch für die Zuordnung der DDB-Objekte zu den »Sachkategorien« (im Folgenden kurz »Kategorisierung«) sollten unterstützend automatische Verfahren eingesetzt werden.

### 5.2.4 Unterstützung der Kategorisierung durch automatische Verfahren

Zwei Drittel der Datensätze in der untersuchten Grundgesamtheit haben keine Werte für eine »Sachkategorie«. Das entspricht hochgerechnet auf die gegenwärtige Gesamtheit der DDB-Objekte (ca. 24 Millionen Anfang 2019) etwa 15,8 Millionen Objekten ohne Zuordnung zu einer Systematikstelle oder Klasse. Es ist nicht anzunehmen, dass die Datengeber eine nachträgliche Kategorisierung ihrer Objekte leisten werden. Deshalb wird eine (semi-)automatische Anreicherung der DDB-Objekte mit Werten für die »Sachkategorie« erforderlich sein.

In einem vierten Schritt wäre daher zu untersuchen, ob die DDB-Objekte automatisch unterstützt dem »Kategoriensystem« zugeordnet werden können. Als Voraussetzung müsste zumindest eines der Elemente »Objektgattung« oder »Themenschlagwort« vorhanden und mit einem Wert belegt sein:

- Kontrollierte Terme sollten aus einem empfohlenen LOD-Vokabular stammen.

---

<sup>211</sup> Das Europeana Vocabulary ist in einem Google-Spreadsheet angelegt. URL: <https://docs.google.com/spreadsheets/d/1qjyyneg6aMoPC2v5hwC8YinmHKNyJtvTJp1HJdnnPc8/>, Dangerfield (2014).

<sup>212</sup> Die Entwicklung der Klassifikation ist im Konferenzpapier »Domain-specific modeling: Towards a Food and Drink Gazetteer« von Tagarev, Tolosi und Alexiev (2015) beschrieben. Das Projekt hat auch einen semantischen Demonstrator entwickelt, siehe <http://efd.ontotext.com/app>.

- Terme ohne Referenz auf ein kontrolliertes Vokabular müssten in einem vorbereitenden Schritt (semi-)automatisch auf Deskriptoren aus den empfohlenen Vokabularen abgebildet werden.

Unter anderen sind folgende Ansätze für eine (semi-)automatische Kategorisierung denkbar:

- Auswertung der Verknüpfungen in den GND-Datensätzen: Unter der Voraussetzung, dass die DDB-Objekte Referenzen zu GND-Deskriptoren enthalten, ließen sich über die Verknüpfungen der GND-Datensätze zur DDC die entsprechenden Klassen der DDC-Sachgruppen ableiten<sup>213</sup>, sodass auch bei fehlender Belegung eines Metadatenelementes für die »Sachkategorie« ein Wert automatisch ermittelt werden kann.
- Auswertung der Verknüpfungen in Wikidata-Datenobjekten: Die Konkordanzen zwischen Wikidata, der Wissensbasis von Wikipedia, und der GND könnten für die Einordnung von DDB-Objekten in Wikipedia-Kategorien ausgewertet werden<sup>214</sup>. Voraussetzung ist auch hier, dass die DDB-Objekte mit GND-Deskriptoren indexiert sind und dass Konkordanzen zwischen Wikipedia-Kategorien und den Kategorien des »Kategoriensystems« bestehen<sup>215</sup>.

Weitere Impulse für den Einsatz automatischer Verfahren könnten von den Ergebnissen der folgenden DFG-Projekte kommen:

- Das Projekt »Aufbau einer Infrastruktur zur Implementierung sachthematischer Zugänge im Archivportal-D am Beispiel Weimarer Republik«<sup>216</sup> setzt auf Verfahren des maschinellen Lernens zur (semi-)automatischen Klassierung von Archivdaten<sup>217</sup>. Inwieweit das zu entwickelnde Verfahren für eine Kategorisierung von DDB-Objekten anderer Sparten übertragbar ist, wäre nach Abschluss der ersten Projektphase im Mai 2020 zu untersuchen<sup>218</sup>.
- Im erwähnten Projekt »GND für Kulturdaten« (GND4C) (siehe Kapitel 5.1, Seite 70) sollen bis April 2020 prototypisch Verfahren eines automatischen Abgleiches unterschiedlicher Erschließungsdaten mit anschließendem maschinell gestützten Mapping zur GND

<sup>213</sup> Die DDC-Sachgruppen können aus den DDC-Deutsch-Klassen durch Trunkierung der Notationen automatisch ermittelt werden, wie in Kapitel 2.4.3 an einem Beispiel illustriert. Als Voraussetzung müssten die DDC-Sachgruppen mit der DDC-Deutsch verlinkt sein und dereferenzierbare URIs haben.

<sup>214</sup> Europeana empfiehlt in einem Blogpost ihren Datenpartnern, ihre Vokabulare mit Wikidata zu verlinken, damit sie von den Einrichtungen der GLAM-Community gemeinsam genutzt werden können, siehe Europeana Foundation (Posted on Monday 7. August 2017).

<sup>215</sup> Zum Beispiel ist der GND-Deskriptor »Ölmalerei« (URI: <http://d-nb.info/gnd/4043238-5>) mit dem Wikidata-Datenobjekt »Ölmalerei« (URI: <https://www.wikidata.org/wiki/Q174705>) verlinkt, der den Kategorien »Technik der Malerei« (URI: <https://www.wikidata.org/wiki/Q8170955>), »Gattung der Malerei« (URI: <https://www.wikidata.org/wiki/Q16862595>) und »Künstlerische Technik« (URI: <https://www.wikidata.org/wiki/Q8803164>) zugeordnet ist.

<sup>216</sup> Laufzeit 1. Projektphase: 1. Juni 2018 - 30. Mai 2020. Eine Beschreibung der Projekthinhalte bietet eine Informationswebsite des Landesarchivs Baden-Württemberg: <https://www.landearchiv-bw.de/web/63528>.

<sup>217</sup> Die Basis sollen Trainingsdaten bieten, die mit einer intellektuell erstellten Klassifikation erschlossen wurden. Das geplante Vorgehen ist in den Arbeitspaketen 2 »Entwicklung einer Referenzsachklassifikation für den Themenbereich „Weimarer Republik“ und intellektuelle Zuordnung von Beständen« und 5 »Definition, Test und Implementierung eines Algorithmus zur Generierung automatischer Zuordnungsempfehlungen« des Projektantrages beschrieben, siehe Hollmann et al. ([2018], S. 7-11).

<sup>218</sup> Die im Rahmen des Projektes zu realisierende Anpassungen an der Datenhaltung und den Datenverarbeitungsprozessen können von der DDB mitbenutzt werden, weil das Archivportal-D auf dem Backend der DDB aufsetzt, vgl. ebenda, S. 5.

entwickelt und erprobt werden<sup>219</sup>. Nach erfolgreich abgeschlossenem Mapping sollen auch Möglichkeiten einer automatischen Anreicherung des Suchindexes der DDB mit Normdaten evaluiert werden<sup>220</sup>.

Entsprechende maschinelle Verfahren setzen jedoch immer ein Minimum an Übereinstimmung der abzugleichenden Daten in wesentlichen Merkmalen voraus. Als Ausgangsdaten für einen automatischen Zeichenkettenvergleich eignen sich in der DDB vorliegende Elementwerte für die vorgeschlagenen Filter daher denkbar schlecht<sup>221</sup>: Fast die Hälfte der untersuchten Indexterme sind nicht gebunden. Das entspricht hochgerechnet auf die Grundgesamtheit aller Indexterme etwa fünf Millionen unkontrollierten Termen in DDB-Objekten (Anfang 2019). Dem Erfolg eines automatischen Abgleiches stehen hier unter anderem Schreibfehler, Synonyme und Homonyme im Weg.

Dass automatische Verfahren die Retrievalqualität in Aggregationsportalen sogar verschlechtern können, zeigen die Ergebnisse semantischer Datenanreicherungen<sup>222</sup> in der Europeana<sup>223</sup>. Die unvermeidliche Konsequenz ist, dass höhere Anforderungen an die Datenqualität gestellt werden müssen<sup>224</sup>.

### 5.2.5 Die Rolle der Datenqualität

»We want better data quality: NOW!« Sechs Jahre nach ihrem Start erhebt die Europeana »Improve Data Quality« zur höchsten strategischen Priorität bis zum Jahr 2020<sup>225</sup>. Ein erster Schritt zu mehr Datenqualität ist das genaue Hinschauen auf die Daten und das Anerkennen der Defizite. Jeni Tennison, Chief Executive Officer am Open Data Institute, beschreibt eindrücklich das Problem der Datenqualität und stellt fest:

If you don't think you have a quality problem with your data, you haven't looked at it yet.  
(Tennison, Thu 6. August 2015)

Von den einzelnen Kultureinrichtungen werden fehlerhafte oder inkomplette Daten weder vollständig aufgedeckt noch korrigiert und aufgearbeitet werden können. Die zukünftige Strategie wird angesichts der Massendaten auf gemeinsames Arbeiten setzen müssen. Anschaulich hat Linux-

---

<sup>219</sup> Das Vorhaben ist im Arbeitspaket »3 Bereitstellung von Schnittstellen und Werkzeugen zur Unterstützung nicht-bibliothekarischer Anwendungskontexte« des GND4C-Projektantrages beschrieben, siehe Niggemann et al. (2017, S. 19-21).

<sup>220</sup> vgl. ebenda S. 21, Task 3.7: Rückfluss von erstellten Konkordanzen in die Fallbeispiele.

<sup>221</sup> Selbst unter günstigen Bedingungen ist ein automatischer Abgleich nur begrenzt erfolgreich, wie Erfahrungen aus dem IN2N-Projekt (Institutionsübergreifende Integration von Normdaten, Laufzeit: Dezember 2012 bis November 2014) für Individualnamen gezeigt haben: Es wurden kontrollierte Personendatensätze aus dem Deutschen Filminstitut DIF e.V. und der GND abgeglichen, vgl. Lindenthal (4. Februar 2016, S. 31); Wiechmann (2015, S. 155).

<sup>222</sup> Europeana bezieht sich mit »semantic enrichment« auf den Prozess, bei dem die Erschließungsdaten der Datengeber automatisch mit LOD-Vokabularen verlinkt werden, indem in den Metadaten vorgefundene Zeichenketten mit kontrollierten Termen der Zielvokabulare abgeglichen werden, vgl. Europeana Foundation (Posted on Thursday 5. November 2015).

<sup>223</sup> Die Anreicherungen gehen vor allem zu Lasten eines präzisen Retrievals, wie z. B. Olensky, Stiller und Dröge (2012, S. 255) in ihrem Fachartikel »Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy« schildern.

<sup>224</sup> Zu diesem Ergebnis kam die EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy in ihrem Abschlussbericht zu semantischen Datenanreicherungen in der Europeana, siehe Stiller, Isaac und Petras (2014, S. 25f.).

<sup>225</sup> vgl. Charles (Posted on Thursday 30. April 2015).

Entwickler Eric S. Raymond ein kollaboratives Vorgehen für die Software-Programmierung beschrieben und es Linus' Gesetz genannt: »Given enough eyeballs, all bugs are shallow.«<sup>226</sup>. Tennison knüpft an die Idee der offenen Plattformen die Hoffnung auf bessere Datenqualität:

The more people look at your data, the more likely they are to find the problems within it. The secret is to build in feedback mechanisms which allow those errors to be corrected, so that you can benefit from those eyes and increase your data quality to what you thought it was in the first place. (Tennison, Thu 6. August 2015)

Wenn diese Vision der vielen Augen vielleicht auch noch zu optimistisch ist, so kann doch im Umkehrschluss gefolgert werden: Ohne kollaborative Plattformen und Strategien für gemeinschaftliches Arbeiten wird die nötige Datenqualität als Voraussetzung für gute Retrievalangebote nicht zu erreichen sein. Die Entwicklung eines gemeinsam genutzten Kategoriensystems wäre ein Anfang.

---

<sup>226</sup> vgl. Raymond (8. August 1999).

# Literatur und Quellen

## Literaturverzeichnis

- BALAKRISHNAN, U. und J.M. AGNE, 2016. *Ergebnisbericht zur Online-Umfrage über den Stand der Sacherschließungsinstrumente (Knowledge Organization Systems, KOS) und der Konkordanzprojekte im deutschsprachigen Raum* [online]. Verfügbar unter: <https://coli-conc.gbv.de/publications/Umfrageergebnisse.pdf>
- BERNERS-LEE, T., March 1989, May 1990. *Information Management: A Proposal* [online]. Verfügbar unter: <https://www.w3.org/History/1989/proposal.html>
- BETZ, F., 2013. *Gattungsbegriffe als Mittel zur Inhaltserschließung alter Drucke. Möglichkeiten und Grenzen* [online]. Masterarbeit. Berlin. Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft. Verfügbar unter: <https://edoc.hu-berlin.de/bitstream/handle/18452/2737/343.pdf>
- BROUGHTON, V. und A. SLAVIC, 2007. Building a faceted classification for the humanities: principles and procedures [online]. *Journal of Documentation*, **63**(5), 727-754 [Zugriff am: 13. Januar 2017]. Verfügbar unter: <https://doi.org/10.1108/00220410710827772>
- CHARLES, V. und C. DEVARENNE, 4. Dezember 2014. *Europeana enriches its data with the AAT* [online]. Verfügbar unter: <https://pro.europeana.eu/page/europeana-aat>
- CHARLES, V., A. ISAAC und T. HILL, 2015. *Discovery - User scenarios and their metadata requirements* [online]. Version 3. Verfügbar unter: [https://pro.europeana.eu/files/Europeana\\_Professional/EuropeanaTech/EuropeanaTech\\_WG/DataQualityCommittee/DQC\\_DiscoveryUserScenarios\\_v3.pdf](https://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_WG/DataQualityCommittee/DQC_DiscoveryUserScenarios_v3.pdf)
- CHARLES, V., Posted on Thursday 30. April 2015. *We want better data quality: NOW!* [online]. Verfügbar unter: <https://pro.europeana.eu/page/data-quality-etech15-roundtables>
- DEUTSCHE FORSCHUNGSGEMEINSCHAFT, 12/2016. *DFG Praxisregeln "Digitalisierung"* [online]. Verfügbar unter: [http://www.dfg.de/formulare/12\\_151/12\\_151\\_de.pdf](http://www.dfg.de/formulare/12_151/12_151_de.pdf)
- DEUTSCHE NATIONABIBLIOTHEK, 2018. *Jahresbericht 2017* [online]. Verfügbar unter: <https://dnb.info/1160486344/34>
- DEUTSCHE NATIONALBIBLIOTHEK, 2017. *Strategische Prioritäten (2017-2020)* [online]. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:101-2017021403>
- EUROPEANA FOUNDATION, 2015. *Europeana Publishing Framework: Metadata* [online]. *Quick Summary*. Verfügbar unter: [https://pro.europeana.eu/files/Europeana\\_Professional/Publications/Publishing\\_Framework/Europeana\\_publishing\\_framework\\_metadata.pdf](https://pro.europeana.eu/files/Europeana_Professional/Publications/Publishing_Framework/Europeana_publishing_framework_metadata.pdf)



- EUROPEANA FOUNDATION, Posted on Monday 7. August 2017. *Why data partners should link their vocabulary to Wikidata: a new case study* [online]. Verfügbar unter: <https://pro.europeana.eu/post/why-data-partners-should-link-their-vocabulary-to-wikidata-a-new-case-study>
- EUROPEANA FOUNDATION, Posted on Thursday 5. November 2015. *Europeana semantic enrichment* [online]. *Find out how Europeana performs semantic enrichment and how you can enrich your metadata with linked open vocabularies.* Verfügbar unter: <https://pro.europeana.eu/page/europeana-semantic-enrichment>
- GOLEM.DE, 2014. *Yahoo schließt seinen Katalog* [online]. Verfügbar unter: <https://www.golem.de/news/yahoo-directory-yahoo-schliesst-seinen-katalog-1409-109478.html>
- GUHA, R., R. MCCOOL und E. MILLER, 2003. *Semantic Search* [online]. Verfügbar unter: <http://www2003.org/cdrom/papers/refereed/p779/ess.html>
- HAFFNER, D. und W. SCHWEIBENZ, 2017. *Ergebnisse der Online-Umfrage zur Nutzung von kontrolliertem Vokabular in Museen im Herbst 2016* [online]. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:bsz:576-opus4-12544>
- HEARST, M., 2009. *Search User Interfaces* [online]. Cambridge: Cambridge University Press. ISBN 9780521113793. Verfügbar unter: <http://searchuserinterfaces.com/book/>
- HILL, T., D. HASKIYA, A. ISAAC, H. MANGUINHAS und V. CHARLES, 2016. *Europeana Search Strategy* [online]. Verfügbar unter: [https://pro.europeana.eu/files/Europeana\\_Professional/Publications/EuropeanaSearchStrategy\\_whitepaper.pdf](https://pro.europeana.eu/files/Europeana_Professional/Publications/EuropeanaSearchStrategy_whitepaper.pdf)
- HOLLMANN, M., MAIER GERALD PROFESSOR DR., E.D. NIGGEMANN und M. RAZUM, [2018]. *Aufbau einer Infrastruktur zur Implementierung sachthematischer Zugänge im Archivportal-D am Beispiel des Themenkomplexes „Weimarer Republik“* [online]. *Antrag auf Förderung eines Vorhabens im Bereich „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS), Förderprogramm „e-Research-Technologien“.* Verfügbar unter: [https://www.landesarchiv-bw.de/sixcms/media.php/120/63271/DFG%20Antrag%20zur%20Implementierung%20sachthematischer%20Zugaenge%20im%20Archivportal\\_D.pdf](https://www.landesarchiv-bw.de/sixcms/media.php/120/63271/DFG%20Antrag%20zur%20Implementierung%20sachthematischer%20Zugaenge%20im%20Archivportal_D.pdf)
- HUBRICH, J., 2018. Semantische Interoperabilität zwischen Klassifikationen und anderen Wissenssystemen. In: H. ALEX, G. BEE und U. JUNGER, Hg. *Klassifikationen in Bibliotheken. Theorie - Anwendung - Nutzen*. Boston: De Gruyter Sauer, S. 267-308.
- KOCH, T. und M. DAY, 1997. *The Role of classification schemes in Internet resource description and discovery* [online]. Specification for resource description methods Part 3. Deliverable 3.2. Verfügbar unter: <http://www.ukoln.ac.uk/metadata/desire/classification/classification.pdf>

- KOCH, T., 2000. Quality-controlled subject gateways: definitions, typologies, empirical overview [online]. *Online Information Review*, **24**(1). Verfügbar unter: <https://doi.org/10.1108/14684520010320040>
- KOCH, T., H. NEUROTH und M. DAY, 2001. *DDC Mapping Report Renardus D7.4* [online]. Verfügbar unter: <http://www.ukoln.ac.uk/metadata/renardus/wp7/ddc-mapping/d74-v01.html>
- KOCH, T., K. GOLUB und A. ARDÖ, 2006. Users Browsing Behaviour in a DDC-Based Web Service: A Log Analysis [online]. *Cataloging & Classification Quarterly*, **42**(3-4), 163-186. Verfügbar unter: [http://dx.doi.org/10.1300/J104v42n03\\_07](http://dx.doi.org/10.1300/J104v42n03_07)
- LEWANDOWSKI, D., 2005. Web Information Retrieval. Technologien zur Informationssuche im Internet [online]. In: DEUTSCHE GESELLSCHAFT FÜR INFORMATIONSWISSENSCHAFT UND INFORMATIONSPRAXIS E.V., DGI, Hg. *Reihe Informationswissenschaft der DGI*. Verfügbar unter: [http://www.bui.haw-hamburg.de/fileadmin/user\\_upload/lewandowski/doc/Web\\_Information\\_Retrieval\\_Buch.pdf](http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Web_Information_Retrieval_Buch.pdf)
- LEWANDOWSKI, D., 2015. *Suchmaschinen verstehen* [online]. Verfügbar unter: <http://dx.doi.org/10.1007/978-3-662-44014-8>
- LINDENTHAL, J., 4. Februar 2016. *Datenqualität und Retrieval* [online]. *Vorschläge zur Verbesserung der Suche in der Deutschen Digitalen Bibliothek*. Verfügbar unter: [http://jl.balilabs.de/DDB/DQ/DDB\\_Datenqualit%C3%A4t\\_Retrieval\\_1.0.pdf](http://jl.balilabs.de/DDB/DQ/DDB_Datenqualit%C3%A4t_Retrieval_1.0.pdf)
- MANGUINHAS, H., 17. November 2016. *Europeana Semantic Enrichment Framework* [online]. *Documentation*. Verfügbar unter: <https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y>
- NIGGEMANN, E., R. GOEBEL, R. KRETZSCHMAR, C. BRACHT und J. AHLERS, 2017. *GND für Kulturdaten (GND4C)* [online]. *1. Projektphase*. Beschreibung des Vorhabens - Projektanträge im Bereich "Wissenschaftliche Literaturversorgungs- und Informationssysteme" (LIS). Verfügbar unter: [https://wiki.dnb.de/download/attachments/134055796/B\\_eRT\\_GND4C\\_public.pdf](https://wiki.dnb.de/download/attachments/134055796/B_eRT_GND4C_public.pdf)
- OLENSKY, M., J. STILLER und E. DRÖGE, 2012. Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy [online]. *Metadata and Semantics Research*, **343**, 252-263. Verfügbar unter: [http://dx.doi.org/10.1007/978-3-642-35233-1\\_25](http://dx.doi.org/10.1007/978-3-642-35233-1_25)
- ONLINE COMPUTER LIBRARY CENTER, 9. Juni 2015. *Change to Dewey Web Services* [online]. Verfügbar unter: <https://www.oclc.org/developer/news/2015/dewey-down.en.html>
- PEKEL, J., 2015a. *Das Museum als Teil des Internets: Ein Interview mit Seb Chan* [online]. Verfügbar unter: <https://www.deutsche-digitale-bibliothek.de/content/journal/background/making-museum-part-internet-interview-seb-chan>

- PEKEL, J., 2015b. *Finden ohne Suche: Ein Interview mit George Oates* [online]. Verfügbar unter: <https://www.deutsche-digitale-bibliothek.de/content/journal/background/discovery-without-search-interview-george-oates>
- RAYMOND, E.S., 8. August 1999. *Die Kathedrale und der Basar* [online]. *SelfLinux-0.12.3*. Verfügbar unter: [https://www.selflinux.org/selflinux/pdf/die\\_kathedrale\\_und\\_der\\_basar.pdf](https://www.selflinux.org/selflinux/pdf/die_kathedrale_und_der_basar.pdf)
- RIVA, P., P. LE BOEUF und M. ZUMER, 2017. *IFLA Library Reference Model* [online]. Dezember 2017. Verfügbar unter: [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017\\_rev201712.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf)
- SCHOLZ, H., 2017. *Europeana Publishing Guide v1.6* [online]. *A guide to the metadata and content requirements for data providers publishing their collections in Europeana*. Verfügbar unter: [https://pro.europeana.eu/files/Europeana\\_Professional/Publications/Europeana%20Publishing%20Guide%20v1.6.pdf](https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Publishing%20Guide%20v1.6.pdf)
- STILLER, J., A. ISAAC und V. PETRAS, 2014. *EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report* [online]. Verfügbar unter: [http://pro.europeana.eu/files/Europeana\\_Professional/EuropeanaTech/EuropeanaTech\\_taskforces/MultilingualSemanticEnrichment//Multilingual%20Semantic%20Enrichment%20report.pdf](http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/MultilingualSemanticEnrichment//Multilingual%20Semantic%20Enrichment%20report.pdf)
- STOCK, W.G. und M. STOCK, 2008. *Wissensrepräsentation* [online]. *Informationen auswerten und bereitstellen*. Verfügbar unter: <http://d-nb.info/1109134258/34>
- TAGAREV, A., L. TOLOSI und V. ALEXIEV, 2015. *Domain-specific modeling: Towards a Food and Drink Gazetteer* [online]. Coimbra. First International Keystone Conference. Verfügbar unter: [https://doi.org/10.1007/978-3-319-59268-8\\_9](https://doi.org/10.1007/978-3-319-59268-8_9)
- TENNISON, J., Thu 6. August 2015. *Five stages of data grief* [online]. Verfügbar unter: <https://theodi.org/article/five-stages-of-data-grief/>
- UMLAUF, K., 1999. Einführung in die bibliothekarische Klassifikationstheorie und -praxis [online]. Mit Übungen. *Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft*, (67). Verfügbar unter: <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h67/>
- VILLA, R., P. CLOUGH, M. HALL und S. RUTTNER, 2013. Search or browse? Casual information access to a cultural heritage collection [online]. *EuroHCIR*, 19-22. Verfügbar unter: <https://pdfs.semanticscholar.org/406a/0f8e1cb1957b4794f357b03118cd8c5388f3.pdf>
- WIECHMANN, B., 2015. Spartenübergreifende Nutzung der GND am Beispiel des Projekts IN2N [online]. *o-bib. Das offene Bibliotheksjournal*, 2(4), 151-160. Verfügbar unter: <https://doi.org/10.5282/o-bib/2015H4S151-160>
- YOUNG, R.D., 2011. *Google Directory Has Been Shut Down* [online]. 25. Juli 2011, 12:00. Verfügbar unter: <https://searchenginewatch.com/sew/news/2096661/google-directory-shut>

## Dokumente der Deutschen Digitalen Bibliothek

FRAUNHOFER IAIS, 21. Mai 2010. *Deutsche Digitale Bibliothek* [online]. *Rahmenbedingungen zur Anforderungsanalyse aus politischer, technischer und funktionaler/technischer Sicht.*

Verfügbar unter: [https://www.iais.fraunhofer.de/content/dam/iais/gf/cts/pdf/DDB-Rahmenbedingungen\\_zur\\_Anforderungsanalyse\\_2010-05-21.pdf](https://www.iais.fraunhofer.de/content/dam/iais/gf/cts/pdf/DDB-Rahmenbedingungen_zur_Anforderungsanalyse_2010-05-21.pdf)

GESCHÄFTSFÜHRUNG DER DEUTSCHEN DIGITALEN BIBLIOTHEK, 2. Mai 2018. *Bericht der Geschäftsführung der DDB zur Sitzung von Vorstand und Mitgliederversammlung am 15.05.2018 in Berlin.* [Unveröffentlichtes Dokument].

SCHULZE, F., 4. Juni 2018. *Datenqualität im Kontext der Deutschen Digitalen Bibliothek* [online].

DDBforum. Verfügbar unter: [http://cms.deutsche-digitale-bibliothek.de/sites/default/files/media/document/2018-06/2018-06-04\\_DDB-Datenqualitaet\\_Schulze.pdf](http://cms.deutsche-digitale-bibliothek.de/sites/default/files/media/document/2018-06/2018-06-04_DDB-Datenqualitaet_Schulze.pdf)

STIFTUNG PREUßISCHER KULTURBESITZ, 2016. *Deutsche Digitale Bibliothek. Kultur und Wissen Online* [online]. *Strategie 2020.* Verfügbar unter: [https://cms.deutsche-digitale-bibliothek.de/sites/default/files/media/documents/ddb\\_strategie\\_2020\\_download.pdf](https://cms.deutsche-digitale-bibliothek.de/sites/default/files/media/documents/ddb_strategie_2020_download.pdf)

STIFTUNG PREUßISCHER KULTURBESITZ, o.J. *Deutsche Digitale Bibliothek* [online]. *Über uns.*

Verfügbar unter: <https://www.deutsche-digitale-bibliothek.de/content/ueber-uns>

TVT CREATIVE MEDIA GMBH IM AUFTRAG DES BEAUFTRAGTEN DER BUNDESREGIERUNG FÜR KULTUR UND MEDIEN, 2012. *Das ist die Deutsche Digitale Bibliothek!* [online]. Verfügbar unter: <https://www.deutsche-digitale-bibliothek.de/content/journal/news/deutsche-digitale-bibliothek>

## Metadatenstandards, Schemata, Anwendungsprofile

ARBEITSGRUPPE ENCODED ARCHIVAL DESCRIPTION DER SOCIETY OF AMERICAN ARCHIVISTS UND DES BÜROS FÜR NETZWERKENTWICKLUNG UND MARC STANDARDS DER LIBRARY OF CONGRESS, 2006. *Encoded Archival Description Tag-Library* [online]. *Version 2002.* Verfügbar unter: <http://www.bundesarchiv.de/imperia/md/content/daofind/1.pdf>

CLAYPHAN, R., 14. Juli 2013. *Europeana Semantic Elements. Specification and Guidelines* [online].

Verfügbar unter:  
[https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/ESE\\_Documentation/Europeana%20Semantic%20Elements%20Specification%20and%20Guidelines%2014%20July%202013.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/ESE_Documentation/Europeana%20Semantic%20Elements%20Specification%20and%20Guidelines%2014%20July%202013.pdf)

COBURN, E., R. LIGHT, G. MCKENNA, R. STEIN und A. VITZTHUM, 2010. *LIDO - Lightweight Information Describing Objects* [online]. *Version 1.0.* 2010-11. Verfügbar unter:

<http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>

- DEUTSCHE NATIONALBIBLIOTHEK, 16. Juni 2014. *Feldbeschreibung der Titeldaten der Deutschen Nationalbibliothek und der Zeitschriftendatenbank im Format MARC 21* [online]. Version 1.6. Verfügbar unter: <http://d-nb.info/1052191681/34>
- DINI-ARBEITSGRUPPE "ELEKTRONISCHES PUBLIZIEREN". *DINI-Zertifikat für Open-Access-Repositorien und -Publikationsdienste 2016* [online]. Version 5.0. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:kobv:11-100239432>
- DLF AQUIFER METADATA WORKING GROUP, 2009. *Digital Library Federation / Aquifer Implementation Guidelines for Sharable MODS Records* [online]. Version 1.1. Verfügbar unter: [https://wiki.dlib.indiana.edu/download/attachments/24288/DLFMODS\\_ImplementationGuidelines.pdf](https://wiki.dlib.indiana.edu/download/attachments/24288/DLFMODS_ImplementationGuidelines.pdf)
- DUBLIN CORE METADATA INITIATIVE, 14. Juni 2012. *Dublin Core Metadata Element Set, Version 1.1* [online]. *Reference Description*. Verfügbar unter: <http://dublincore.org/documents/2012/06/14/dces/>
- FACHSTELLE MUSEUM DER DEUTSCHEN DIGITALEN BIBLIOTHEK IM INSTITUT FÜR MUSEUMSFORSCHUNG, 2014. *Globale Vorgaben für DDB-LIDO* [online]. Verfügbar unter: [https://www.smb.museum/fileadmin/website/Institute/Institut\\_fuer\\_Museumsforschung/Fachstelle\\_Museum/Handreichung\\_DDB-LIDO\\_1.0.pdf](https://www.smb.museum/fileadmin/website/Institute/Institut_fuer_Museumsforschung/Fachstelle_Museum/Handreichung_DDB-LIDO_1.0.pdf)
- FICHTL, B., M. FREIBERG, A. KAILUS, G. KNAUS und R. STEIN, 2015. *Gemeinsamer Feldkatalog Graphischer Sammlungen* [online]. Version 1.0. Verfügbar unter: [https://www.arthistoricum.net/fileadmin/groups/arthistoricum/Netzwerke/Graphik\\_vernetzt/Feldkatalog\\_Graphik\\_1.0.pdf](https://www.arthistoricum.net/fileadmin/groups/arthistoricum/Netzwerke/Graphik_vernetzt/Feldkatalog_Graphik_1.0.pdf)
- J. PAUL GETTY TRUST und ARTSTOR, 17. Juli 2006. *CDWA Lite* [online]. *Specification for an XML Schema Specification for an XML Schema for Contributing Records via the OAI Harvesting Protocol*. Verfügbar unter: [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/cdwalite.pdf](http://www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.pdf)
- KOMPETENZZENTRUM INTEROPERABLE METADATEN, 22. August 2007. *Deutsche Übersetzung des Dublin-Core-Metadata-Elemente-Sets, Version 1.1* [online]. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:101:1-200911103125>
- KRAUTH, W., 8. Januar 2013. *EAD(DDB)-Profil basierend auf EAD Version 2002* [online]. Verfügbar unter: [https://www.landesarchiv-bw.de/sixcms/media.php/120/55627/Excel\\_EAD\\_DDB\\_1.1\\_Profil.xls](https://www.landesarchiv-bw.de/sixcms/media.php/120/55627/Excel_EAD_DDB_1.1_Profil.xls)
- LANDESARCHIV BADEN-WÜRTTEMBERG, 2013. *EAD(DDB) 1.1* [online]. *Content von Archiven: Feldinhalte und EAD-Elemente der Findbuch-EAD*. Verfügbar unter: [https://www.landesarchiv-bw.de/sixcms/media.php/120/55625/EAD-DDB\\_1.1\\_Findbuch\\_EAD.pdf](https://www.landesarchiv-bw.de/sixcms/media.php/120/55625/EAD-DDB_1.1_Findbuch_EAD.pdf)

- LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 1999.  
*MARC 21 Format for Bibliographic Data* [online]. Update No. 25 (December 2017). Verfügbar unter: <http://www.loc.gov/marc/bibliographic/>
- LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 8. Januar 2018. *MODS User Guidelines (Version 3)* [online]. Verfügbar unter: <https://www.loc.gov/standards/mods/userguide/index.html>
- MCKENNA, G. und R. STEIN, 2015. *Implementing LIDO* [online]. Verfügbar unter: <http://www.athenaplus.eu/getFile.php?id=559>
- RÜHLE, S., A. JAHNKE und G. KÜHLE, 2017. *MODS Anwendungsprofil für digitalisierte Medien* [online]. Version 2.3. Verfügbar unter: [https://dfg-viewer.de/fileadmin/groups/dfgviewer/MODS-Anwendungsprofil\\_2.3.pdf](https://dfg-viewer.de/fileadmin/groups/dfgviewer/MODS-Anwendungsprofil_2.3.pdf)
- STEIN, R., 26. September 2011. *Introduction to LIDO* [online]. Rom. Tutorial for the Linked Heritage Project. Verfügbar unter: <http://www.linkedheritage.eu/getFile.php?id=180>
- TECHNICAL SUBCOMMITTEE FOR ENCODED ARCHIVAL STANDARDS OF THE SOCIETY OF AMERICAN ARCHIVISTS und LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 2016. *Encoded Archival Description Tag Library Version EAD3* [online]. Verfügbar unter: <http://www.loc.gov/ead/EAD3taglib/index.html>
- THE J. PAUL GETTY TRUST, 24. Oktober 2017. *Metadata Standards Crosswalk* [online]. Verfügbar unter: [http://www.getty.edu/research/publications/electronic\\_publications/intrometadata/crosswalks.html](http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html)

## Vokabularstandards, Regelwerke und Anwendungsleitfäden zur Indexierung

- DEUTSCHE NATIONALBIBLIOTHEK, 1. März 2017. *EH-S-01* [online]. *Identifizierender Zusatz bei Sachbegriffen*. Verfügbar unter: <https://wiki.dnb.de/download/attachments/116426796/EH-S-01.pdf>
- DEUTSCHE NATIONALBIBLIOTHEK, 1. Juni 2014. *DDC-Sachgruppen der deutschen Nationalbibliografien* [online]. *Leitfaden zu ihrer Vergabe*. Version 1.1. Verfügbar unter: <https://d-nb.info/1052700705/34>
- DEUTSCHE NATIONALBIBLIOTHEK, 11. November 2011. *GND-Systematik* [online]. *Leitfaden zu ihrer Vergabe*. Version 1.1. Verfügbar unter: <https://d-nb.info/1018626042/34>
- DEUTSCHE NATIONALBIBLIOTHEK, ARBEITSSTELLE FÜR STANDARDISIERUNG, März 2017. *Regeln für die Schlagwortkatalogisierung* [online]. *RSWK*. 4. , vollständig überarbeitete Auflage. Verfügbar unter: <https://d-nb.info/1126513032/34>

Deutsches Institut für Normung e.V. DIN 1463-1:1987-11-00, *Erstellung und Weiterentwicklung von Thesauri*.

Deutsches Institut für Normung e.V. DIN 32705:1987, *Klassifikationssysteme. Erstellung und Weiterentwicklung von Klassifikationssystemen*.

FACHHOCHSCHULE KÖLN, 30. September 2010. *Leitfaden zur Vergabe von DDC-Notationen an SWD-Schlagwörtern* [online]. Verfügbar unter: [https://ixtrieve.fh-koeln.de/crisscross/CrissCross\\_Endg\\_Grundlagenpapier\\_Sept2010.pdf](https://ixtrieve.fh-koeln.de/crisscross/CrissCross_Endg_Grundlagenpapier_Sept2010.pdf)

International Standard Office. ISO 25964-2:2013, *Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies*.

International Standard Office. ISO 25964-1:2011, *Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval*.

SOCIETY OF AMERICAN ARCHIVISTS, 2013. *Describing Archives* [online]. A content standard. Verfügbar unter: [http://files.archivists.org/pubs/DACS2E-2013\\_v0315.pdf](http://files.archivists.org/pubs/DACS2E-2013_v0315.pdf)

THE J. PAUL GETTY TRUST, 23. Oktober 2017. *Categories for the Description of Works of Art* [online]. Verfügbar unter: [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/index.html](http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html)

VERBUNDZENTRALE DES GEMEINSAMEN BIBLIOTHEKSVERBUNDES, 19 Juni 2013, 12:00. *Alte Drucke* [online]. Verfügbar unter: <http://www.gbv.de/du/katricht/sondadr.pdf>

VISUAL RESOURCES ASSOCIATION, 2006. *CCO Commons* [online]. *Cataloguing Cultural Objects*. Verfügbar unter: <http://cco.vrafoundation.org/>

## Einzelne Vokabulare, Schlagwortlisten, Codelisten

ALEX, H., 16. Mai 2018. *FAQ zur DDC allgemein und zur Anwendung der DDC Deutsch* [online]. Verfügbar unter: [http://www.ddc-deutsch.de/Subsites/ddcdeutsch/DE/FAQ/faq\\_node.html](http://www.ddc-deutsch.de/Subsites/ddcdeutsch/DE/FAQ/faq_node.html)

ARBEITSGEMEINSCHAFT ALTE DRUCKE BEIM GBV, 11. Juli 2011. *AAD Gattungs- und Sachbegriffe* [online]. Verfügbar unter: [https://verbundwiki.gbv.de/download/attachments/52592772/aad\\_gattung.pdf](https://verbundwiki.gbv.de/download/attachments/52592772/aad_gattung.pdf)

DANGERFIELD, M.C., 2014. *Europeana Vocabulary Final Draft* [online]. Verfügbar unter: <https://docs.google.com/spreadsheets/d/1qjyyneg6aMoPC2v5hwC8YinmHKNyJtvTJp1HJdnnPc8/>

DCMI USAGE BOARD, 14. Juni 2012. *DCMI Type Vocabulary* [online]. Verfügbar unter: <http://dublincore.org/documents/2012/06/14/dcmi-terms/>

DEUTSCHE FORSCHUNGSGEMEINSCHAFT, 2015. *DFG-Vordruck 12.109 – 02/15* [online]. *Richtlinien zur überregionalen Literaturversorgung der Sondersammelgebiete und virtuellen Fachbibliotheken*. Verfügbar unter: [http://www.dfg.de/formulare/12\\_109/12\\_109\\_de.pdf](http://www.dfg.de/formulare/12_109/12_109_de.pdf)

DEUTSCHE NATIONALBIBLIOTHEK, 1. Januar 2011. *DDC-Sachgruppen ab Bibliografie-Jahrgang 2004 für den Neuerscheinungsdienst und die Reihen A, B, C, H und O der Deutschen Nationalbibliografie mit enthaltenen DDC-Klassen* [online]. Auf Grundlage der 23. Ausgabe der DDC. Verfügbar unter:

<http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/ddcSachgruppenDNB.pdf>

DEUTSCHE NATIONALBIBLIOTHEK, 25. Juli 2012. *Liste der in der Deutschen Nationalbibliothek verwendeten Gattungsbegriffe* [online]. Verfügbar unter:

[http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/listeGattungsbegriffe.pdf?\\_\\_blob=publicationFile](http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/listeGattungsbegriffe.pdf?__blob=publicationFile)

DEUTSCHE NATIONALBIBLIOTHEK, ARBEITSSTELLE FÜR STANDARDISIERUNG, 28. November 2017. *Normierte Begriffe zur Beschreibung des Inhalts RDA 7.2.1.3, D-A-CH* [online]. *Arbeitshilfe*. Version 1.8. Verfügbar unter: <https://wiki.dnb.de/download/attachments/106042227/AH-007.pdf>

DEUTSCHES DOKUMENTATIONSZENTRUM FÜR KUNSTGESCHICHTE - BILDARCHIV FOTO

MARBURG, [o. J.]. *Partage Plus all concepts* [online]. Verfügbar unter:

<http://partage.vocnet.org/html/>

DINI-ARBEITSGRUPPE "ELEKTRONISCHES PUBLIZIEREN", 2010. *Gemeinsames Vokabular für Publikations- und Dokumenttypen* [online]. Verfügbar unter: <https://doi.org/10.18452/1492>

GEMEINSAMER BIBLIOTHEKSVERBUND, 9. August 2018. *Gattungsbegriffe der Arbeitsgemeinschaft Alte Drucke beim GBV und SWB* [online]. *GBV Verbundwiki*. Verfügbar unter:

<https://verbundwiki.gbv.de/x/DwBpB>

INITIATIVE MUSEUMSVOKABULAR.DE, 2007. *Vokabulare in der Museumsdokumentation* [online]. Verfügbar unter:

<http://museum.zib.de/museumsvokabular/index.php?main=home&ls=9&co=we&ln=de>

LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 13. April 2018. *Classification Scheme Source Codes* [online]. Verfügbar unter:

<https://www.loc.gov/standards/sourcelist/classification.html>

LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 20. Oktober 2017. *MARC Genre Term List* [online]. Verfügbar unter:

<https://www.loc.gov/standards/valuelist/marcgt.html>

LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 7. September 2018. *Subject Heading and Term Source Codes* [online]. Verfügbar unter:

<https://www.loc.gov/standards/sourcelist/subject.html>

LIBRARY OF CONGRESS - NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE, 8 Juni 2018, 12:00. *Genre/Form Code and Term Source Codes* [online]. Verfügbar unter:

<https://www.loc.gov/standards/sourcelist/genre-form.html>



PAGEL, C. und R. STEIN, 2013. *Creation of a specialist art nouveau terms "micro" vocabulary* [online]. D3.2. Verfügbar unter: <http://www.partage-plus.eu/en/servefile?id=56>

RIJKSBUROU VOOR KUNSTHISTORISCHE DOCUMENTATIE, 2012. *Iconclass* [online]. Verfügbar unter: <http://www.iconclass.nl/home>

THE J. PAUL GETTY TRUST, 20. August 2018. *About the AAT* [online]. Verfügbar unter: <http://www.getty.edu/research/tools/vocabularies/aat/about.html#purpose>

UNIVERSITÄTSBIBLIOTHEK REGENSBURG, 2018. *Regensburger Verbundklassifikation Online* [online]. Verfügbar unter: <https://rvk.uni-regensburg.de/regensburger-verbundklassifikation-online>

Alle Hyperlinks in dieser Arbeit wurden zuletzt am 05.10.2020 geprüft.

## Forschungsdatenverzeichnis

Alle hier aufgeführten ZIP-Dateien befinden sich als Forschungsdokumentation auf dem edoc-Server der Humboldt-Universität zu Berlin. Sie können dort über den Digital Objekt Identifier <https://doi.org/10.18452/19655> aufgerufen werden.

### **Programmierskripte zur Ermittlung der DDB-Stichprobe, Datei »ddb-process.zip«**

Diese ZIP-Datei enthält Skripte in der Programmiersprache PHP, die zum Zweck der Ermittlung einer randomisierten Stichprobe aus der Grundgesamtheit aller Datensätze in der DDB, die Objekte des kulturellen Erbes beschreiben, erstellt wurden. Die Stichprobe wurde im Februar 2017 durch eine Abfrage der offenen Programmierschnittstelle der DDB (Application Programming Interface) ermittelt. Diese bildet die Grundlage für die Datenanalyse, die im Rahmen dieser Arbeit angefertigt wurde.

### **Datenbankabzug der DDB-Stichprobe, Datei »ddb-sample.sql.zip«**

Diese ZIP-Datei beinhaltet den Datenabzug aus der MySQL-Datenbank mit 20.000 objektbeschreibenden Datensätzen der DDB, die vom 20. bis 22. Februar 2017 anhand einer Zufallsvariable aus der Grundgesamtheit von 20,5 Millionen Datensätzen in der DDB gezogen wurden.

### **Aussagekraft der DDB-Stichprobe, Datei »ddb-stichprobe-aussagekraft.zip«**

Diese ZIP-Datei umfasst zwei Arbeitsmappen in den Formaten CSV und XSLX, die der Überprüfung der DDB-Stichprobe und der Berechnung ihrer Aussagekraft dienen.

In der Arbeitsmappe »ddb-stichprobe« wurden Metainformationen zur Stichprobe erfasst und anhand einer Merkmalsmatrix aus Metadatenformat und Datengeber in 201 Datensets angeordnet.

Die Arbeitsmappe »ddb-provider-datasets« enthält Berechnungen zur Verteilung der Metadatenformate auf die DDB-Objekte in der Stichprobe und in der Grundgesamtheit. Die Vergleichszahlen für die Grundgesamtheit stammen aus einer Erhebung der Servicestelle DDB vom 11. April bis zum 07. Mai 2018.

### **Auswertung der DDB-Stichprobe, Datei »ddb-datenauswertung.zip«**

Diese ZIP-Datei enthält 20 Arbeitsmappen, jeweils in den Formaten CSV und XSLX, mit den Auswertungen der Werte in den untersuchten Metadatenelementen.

Für jedes Metadatenformat gibt es eine Arbeitsmappe mit einer Haupttabelle, in der alle aus den Untersuchungsdatensätzen ausgelesenen Elementwerte und Attributwerte in einer Zeile angeordnet sind: »marc\_stichprobe«, »mods\_stichprobe«, »ead\_stichprobe«, »lido\_stichprobe« und »dc\_stichprobe«.

Pro Metadatenformat gibt es wiederum für jedes Untersuchungselement eine Arbeitsmappe mit den Auswertungen der Werte. Die Namen der Arbeitsmappen bestehen aus dem Kürzel des

Metadatenstandards und dem Elementnamen. Für das Metadatenformat MACHine-Readable Cataloging (MARC) wurden Arbeitsmappen angelegt:

- »marc\_82« (»Notation nach der Dewey Decimal Classification«),
- »marc\_83« (»Zusätzliche DDC-Notation«),
- »marc\_84« (»Andere Notation«),
- »marc\_650« (»Nebeneintragung unter einem Schlagwort - Sachschlagwort«),
- »marc\_655« (»Indexierungsterm - Genre/Formschlagwort«).

Folgende Arbeitsmappen enthalten die Auswertungen der Elementwerte für das Metadatenformat Metadata Object Description Schema (MODS):

- »mods\_classification« (»Klassifikation«),
- »mods\_topic« (»Thema«),
- »mods\_genre« (»Gattungsbegriff«).

Folgende Arbeitsmappen enthalten die Auswertungen der Elementwerte für das Metadatenformat Encoded Archival Description (EAD):

- »ead\_subject« (»Gegenstand«),
- »ead\_genreform« (»Genre- und Formbegriffe«).

Folgende Arbeitsmappen enthalten die Auswertungen der Elementwerte für das Metadatenformat Lightweight Information Describing Objects (LIDO):

- »lido\_classification« (»Classification«),
- »lido\_subjectConcept« (»SubjectConcept«),
- »lido\_objectWorkType« (»Object/Work Type«).

Folgende Arbeitsmappen enthalten die Auswertungen der Elementwerte für das Dublin Core Metadata Elements Set (DCMES):

- »dcmes\_subject« (»Thema«),
- »dcmes\_type« (»Typ«).

## Anhang A: Zusammenstellung der untersuchten Metadatenelemente

Die folgenden Tabellen führen alle Elemente und Attribute aus den Metadatenschemata und Anwendungsprofilen auf, die bei der Untersuchung der Lieferdaten der DDB berücksichtigt wurden. Die Schreibweisen der Elementnamen in der Tabellenzeile »Bezeichnung« wurden zur besseren Lesbarkeit vereinheitlicht. Die Kardinalität, also der Verpflichtungsgrad und die Wiederholbarkeit eines Elementes, ist – sofern nicht anders vermerkt – in der Chen-Notation ausgedrückt. Die Definitionen und Kommentare stammen aus den Spezifikationen der Metadatenschemata und begleitenden Anwenderleitfäden. Die Konkordanzen der Elemente aus den verschiedenen Standards zueinander basieren auf dem »Metadata Standards Crosswalk« des Getty Research Institute<sup>227</sup>, den Dokumentationen der Standards und eigenen Ergänzungen der Autorin. In der letzten Tabellenzeile ist das jeweilige Element dem vorgeschlagenen Informationsfilter der DDB zugeordnet.

### MAchine-Readable Cataloging, 1999 Edition

Für die XML-Variante des MARC-Formates für bibliografische Daten wird in dieser Arbeit die Schreibweise von MARC 21 benutzt. Die Nummern der Felder stehen stellvertretend für Elemente des MARCXML-Schemas. Das MARC 21-Feld 082 beispielsweise wird in MARCXML so formatiert: `<datafield tag="082">`. Die mit dem Zeichen \$ eingeleiteten Nummern stehen in MARC 21 für Unterfelder und in MARCXML für Unterelemente, zum Beispiel `<subfield code="a">`.

Das MARCXML-Schema macht keine Aussagen zur Kardinalität der Elemente. Die Angaben zur Wiederholbarkeit der Felder sind aus der »Feldbeschreibung der Titeldaten der Deutschen Nationalbibliothek und der Zeitschriftendatenbank im Format MARC 21« entnommen. Dabei steht (NR) für not repeatable/nicht wiederholbar und (R) für repeatable/wiederholbar.

Feld	082
Bezeichnung	Dewey Decimal Classification Number (englisch) ; Notation nach der Dewey Decimal Classification (deutsch)
URI	<a href="http://www.loc.gov/marc/bibliographic/bd082.html">http://www.loc.gov/marc/bibliographic/bd082.html</a>
Kardinalität	(NR)
Definition	»Classification number is taken from Dewey Decimal Classification and Relative Index [...]« (Library of Congress - Network Development and MARC Standards Office, 1999, <a href="http://www.loc.gov/marc/bibliographic/bd082.html">http://www.loc.gov/marc/bibliographic/bd082.html</a> )
Kommentar	Inhalte/Erläuterungen für das Unterfeld \$a (R) Classification Number (Notation): »DDC-Hauptsachgruppe (Übersicht der Sachgruppen siehe: <a href="http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html">http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html</a> ) [...] DDC-Nebensachgruppe (Übersicht der Sachgruppen siehe:

<sup>227</sup> vgl. The J. Paul Getty Trust (24. Oktober 2017).

	<a href="http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html">http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html</a> ) [...] DDC-Notation« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 77–78)  Inhalte/Erläuterungen für das Unterfeld \$2 (NR) Edition Number (Ausgabennummer): »„22sdb“ – DDC-Sachgruppen, basierend auf DDC-Ausgabe 22 „23sdb“ – DDC-Sachgruppen, basierend auf DDC-Ausgabe 23 „22/ger“ – Deutsche DDC-Ausgabe 22 „22/eng“ – Englische DDC-Ausgabe 23 „DDCbio01“ – BioDDC (Ausgabe 2007)« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 79–80)
Unterfelder	\$a Classification Number (englisch) , Notation (deutsch) ; \$2 Edition Number (englisch) , Ausgabennummer (deutsch)
Konkordanzen	MODS <classification> ; LIDO <classification> ; DCMES <subject>
Filter	Sachkategorie

<b>Feld</b>	083
Bezeichnung	Additional Dewey Decimal Classification Number (englisch) ; Zusätzliche DDC-Notation (deutsch)
URI	<a href="http://www.loc.gov/marc/bibliographic/bd083.html">http://www.loc.gov/marc/bibliographic/bd083.html</a>
Kardinalität	(R)
Definition	»Classification number used for subject access. It is taken from Dewey Decimal Classification and Relative Index [...]« (Library of Congress - Network Development and MARC Standards Office, 1999, <a href="http://www.loc.gov/marc/bibliographic/bd083.html">http://www.loc.gov/marc/bibliographic/bd083.html</a> )
Kommentar	Inhalte/Erläuterungen für das Unterfeld \$a (R) Classification Number (Notation): »wie 082« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 81–82)  Inhalte/Erläuterungen für das Unterfeld \$2 (NR) Edition Number (Ausgabennummer): »wie 082« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 81–82)
Unterfelder	\$a Classification Number (englisch) , Notation (deutsch) ; \$2 Edition Number (englisch) , Ausgabennummer (deutsch)
Konkordanzen	MODS <classification> ; LIDO <classification> ; DCMES <subject>
Filter	Sachkategorie

<b>Feld</b>	084
Bezeichnung	Other Classification Number (englisch) ; Andere Notation (deutsch)
URI	<a href="http://www.loc.gov/marc/bibliographic/bd084.html">http://www.loc.gov/marc/bibliographic/bd084.html</a>
Kardinalität	(R)
Definition	»Classification number from a scheme not covered by one of the other number fields.« (Library of Congress - Network Development and MARC Standards Office, 1999, <a href="http://www.loc.gov/marc/bibliographic/bd084.html">http://www.loc.gov/marc/bibliographic/bd084.html</a> )

Kommentar	<p>Inhalte/Erläuterungen für das Unterfeld \$a (R) Classification Number (Notation):</p> <p>»DDC-Hauptsachgruppe (Übersicht der Sachkategorien siehe: <a href="http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html">http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html</a>) [...]</p> <p>DDC-Nebensachgruppe (Übersicht der Sachkategorien siehe: <a href="http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html">http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html</a>) [...]</p> <p>SWD- bzw. GND-Systematiknummer (Angaben zur 1. bis 10. Kette) [...]</p> <p>Nummern der Sondersammelgebiete nach dem DFG-Sondersammelgebietsplan« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 84–85)</p> <p>Inhalte/Erläuterungen für das Unterfeld \$2 (NR) Source of Number (Quelle der Nummer):</p> <p>»“sdbn” – DDC-Sachgruppen der DNB</p> <p>„sswd“ – SWD- bzw. GND-Systematik</p> <p>„ssgn“ – Nummern der Sondersammelgebiete« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 86)</p>
Unterfelder	\$a Classification Number (englisch) , Notation (deutsch) ; \$2 Source of Number , Ausgabennummer
Konkordanzen	MODS <classification> ; LIDO <classification> ; DCMES <subject>
Filter	Sachkategorie

Feld	655
Bezeichnung	Index Term - Genre/Form (englisch) ; Indexierungsterm - Genre/Formschlagwort (deutsch)
URI	<a href="http://www.loc.gov/marc/bibliographic/bd655.html">http://www.loc.gov/marc/bibliographic/bd655.html</a>
Kardinalität	(R)
Definition	<p>»Terms indicating the genre, form, and/or physical characteristics of the materials being described. A genre term designates the style or technique of the intellectual content of textual materials or, for graphic materials, aspects such as vantage point, intended purpose, characteristics of the creator, publication status, or method of representation. A form term designates historically and functionally specific kinds of materials distinguished by their physical character, the subject of their intellectual content, or the order of information within them. Physical characteristic terms designate historically and functionally specific kinds of materials as distinguished by an examination of their physical character, subject of their intellectual content, or the order of information with them.</p> <p>Examples of genre terms for textual materials are: biographies, catechisms, essays, hymns, or reviews. Examples of form and physical characteristic terms are: daybooks, diaries, directories, journals, memoranda, questionnaires, syllabi, or time sheets.</p> <p>May contain terms constructed from a faceted vocabulary. For each term found in the field, an identification is given as to the facet/hierarchy in the thesaurus from which the term came. In addition, identification is given as to which term is the focus term of the expression.</p> <p>Standard published lists are used for the genre and form terms. The lists are identified either by the second indicator (Thesaurus) or by subfield \$2 (Source of term).« (Library of Congress - Network Development and MARC Standards Office, 1999, <a href="http://www.loc.gov/marc/bibliographic/bd655.html">http://www.loc.gov/marc/bibliographic/bd655.html</a>)</p>

Kommentar	<p>Inhalte/Erläuterungen für das Unterfeld \$0 (R) Authority Record Control Number (IDN des Normdatensatzes):</p> <p>»“(DE-101)”IDN des Normdatensatzes Literarische Gattung (Satzart Tk)« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 228)</p> <p>Inhalte/Erläuterungen für das Unterfeld \$a (R) Genre/Form Data or Focus Term (Genre/Form oder fokussierter Term):</p> <p>»Literarische Gattungsbegriffe (Liste s. <a href="http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/listeGattungsbegriffe.pdf">http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/listeGattungsbegriffe.pdf</a>)« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 227)</p> <p>Inhalte/Erläuterungen für das Unterfeld \$2 (R) Source of Term (Quelle des Terms):</p> <p>»“gnd“ [...] „gatbeg“ – Literarischer Gattungsbegriff« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 228)</p>
Unterfelder	\$0 Authority Record Control Number , IDN des Normdatensatzes ; \$a Genre/Form Data or Focus Term , Genre/Form oder fokussierter Term ; \$2 Source of Term , Quelle des Terms
Konkordanzen	MODS <genre> ; EAD <genreform> ; LIDO <objectWorkType> ; DCMES <type>
Filter	Objektgattung

<b>Feld</b>	650
Bezeichnung	Subject Added Entry – Topical Term (englisch) ; Nebeneintragung unter einem Schlagwort - Sachschlagwort (deutsch)
URI	<a href="http://www.loc.gov/marc/bibliographic/bd650.html">http://www.loc.gov/marc/bibliographic/bd650.html</a>
Kardinalität	(R)
Definition	<p>»Subject added entry in which the entry element is a topical term.</p> <p>Topical subject added entries may consist of general subject terms including names of events or objects. Subject added entries are assigned to a bibliographic record to provide access according to generally accepted thesaurus-building rules (e.g., Library of Congress Subject Headings (LCSH), Medical Subject Headings (MeSH). Field 650 may be used by any institution assigning subject headings based on the lists and authority files identified in the second indicator position or in subfield \$2 (Source of heading or term).</p> <p>A title (e.g., Bible and atheism), a geographic name (e.g., Iran in the Koran), or the name of a corporate body (e.g., Catholic Church and humanism) used in a phrase subject heading are also recorded in field 650.« (Library of Congress - Network Development and MARC Standards Office, 1999, <a href="http://www.loc.gov/marc/bibliographic/bd650.html">http://www.loc.gov/marc/bibliographic/bd650.html</a>)</p>
Kommentar	<p>Inhalte/Erläuterungen für das Unterfeld \$0 (R) Authority Record Control Number (IDN des Normdatensatzes):</p> <p>»“(DE-101)”IDN des Normdatensatzes [...]</p> <p>„(DE-588)”IDN des Normdatensatzes« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 219)</p> <p>Inhalte/Erläuterungen für das Unterfeld &lt;\$a&gt; (R) Topical Term or Geographic Name as Entry Element (Sachschlagwort oder geografischer Name als Eintragungselement):</p> <p>»Schlagwörter aus Altdaten der Deutschen Nationalbibliothek: Hauptschlagwort [...]« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 218)</p>

	Inhalte/Erläuterungen für das Unterfeld \$2 (R) Source of Heading or Term (Quelle der Ansetzung oder des Terms) »“gnd“ [...] „local“« (Deutsche Nationalbibliothek, 16. Juni 2014, S. 219–220)
Unterfelder	\$0 Authority Record Control Number , IDN des Normdatensatzes ; \$a Topical Term or Geographic Name as Entry Element , Sachschlagwort oder geografischer Name als Eintragungselement ; \$2 Source of Heading or Term , Quelle der Ansetzung oder des Terms
Konkordanzen	MODS <topic> ; EAD <subject> ; LIDO <subjectConcept> ; DCMES <subject>
Filter	Themenschlagwort

## Metadata Object Description Schema, MODS 3.6

<b>Element</b>	<classification>
Bezeichnung	Classification (englisch) ; Klassifikation (deutsch)
URI	<a href="http://www.loc.gov/standards/mods/mods-outline-3-6.html#classification">http://www.loc.gov/standards/mods/mods-outline-3-6.html#classification</a>
Kardinalität	0..n
Definition	»A designation applied to a resource that indicates the subject by applying a formal system of coding and organizing resources according to subject areas.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/classification.html">https://www.loc.gov/standards/mods/userguide/classification.html</a> )
Kommentar	<p>»&lt;classification&gt; contains the classification or call number of a resource. [...]</p> <p>The DLF/Aquifer Implementation Guidelines for Shareable MODS Records state that the use of the &lt;classification&gt; element is optional.</p> <p>The DLF/Aquifer guidelines recommend that &lt;classification&gt; contain only classification numbers and call numbers whose authorities are referenced in Classification Scheme Source Codes [<a href="https://www.loc.gov/standards/sourcelist/classification.html">https://www.loc.gov/standards/sourcelist/classification.html</a>] maintained by the Library of Congress. It is left to the institution's discretion whether to truncate assigned call numbers to just the formal classification segment (for example, QA76.17), or to include the full call number (for example, QA76.17.T55).</p> <p>The authority attribute is required. All occurrences of the MODS &lt;classification&gt; element should contain the attribute authority to indicate the name of the classification scheme used in the element. Values for this attribute should come from the Library of Congress Classification Scheme Source Codes [<a href="https://www.loc.gov/standards/sourcelist/classification.html">https://www.loc.gov/standards/sourcelist/classification.html</a>].« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/classification.html">https://www.loc.gov/standards/mods/userguide/classification.html</a>)</p> <p>»Für Klassifikationen gilt:</p> <p>Der Name der Klassifikation wird im Attribut authority angegeben, bzw. durch einen URI im Attribut authorityURI referenziert. Die Notation kann über eine Angabe in valueURI referenziert werden. Ist dies der Fall, so kann auf das Attribut authority verzichtet werden. [...]</p>



	<p>Eines der folgenden Attribute muss verwendet werden:</p> <p>authority: Name der verwendeten Klassifikation</p> <p>authorityURI: URI der verwendeten Klassifikation (z. B. <a href="http://dewey.info">http://dewey.info</a>)</p> <p>valueURI: URI der Notation.« (Rühle, Jahnke und Kühle, 2017, S. 21)</p>
Attribute	authority; authorityURI; valueURI
Konkordanzen	MARC 082 ; MARC 083 ; MARC 084 ; LIDO <classification> ; DCMES <subject>
Filter	Sachkategorie

Element	<genre>
Bezeichnung	Genre (englisch) ; Gattungsbegriff (deutsch)
URI	<a href="http://www.loc.gov/standards/mods/mods-outline-3-6.html#genre">http://www.loc.gov/standards/mods/mods-outline-3-6.html#genre</a>
Kardinalität	0..n
Definition	<p>»A term(s) that designates a category characterizing a particular style, form, or content, such as artistic, musical, literary composition, etc.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/genre.html">https://www.loc.gov/standards/mods/userguide/genre.html</a>)</p>
Kommentar	<p>»&lt;genre&gt; contains terms that give more specificity for the form of an object than the broad terms used in &lt;typeOfResource&gt; [<a href="https://www.loc.gov/standards/mods/userguide/dlfaquifer.html#typeofresource">https://www.loc.gov/standards/mods/userguide/dlfaquifer.html#typeofresource</a>]. The terms may be from a controlled list with a designation of the authoritative list used in the authority attribute, or it may be an uncontrolled term. If no authority is specified, it is assumed that the term is uncontrolled. For an example genre list, see MARC Genre Term List [<a href="https://www.loc.gov/standards/valuelist/marctgt.html">https://www.loc.gov/standards/valuelist/marctgt.html</a>].</p> <p>The &lt;genre&gt; element should be used to characterize the content of the resource rather than the resource itself which would be &lt;form&gt;. [...]</p> <p>The DLF /Aquifer Implementation Guidelines for Shareable MODS Records recommend the use of at least one &lt;genre&gt; element in every MODS record and, if a value is provided, require the use of a value from a controlled list and the designation of this list in the authority attribute.</p> <p>Genre is a term that carries different specific meanings within different communities of practice, and the content of most information objects can be characterized by genre at some level of granularity, either very broad or quite specific. For example, photographs and ambrotypes are both valid genre characterizations, depending on your perspective, for a specific type of direct positive photographic print. At the very least, institutions can provide a very broad genre term for materials being digitized. Broad terms appear in many standard thesauri used for genre terms. For example, the genres "books" and "photographs" appear in the Library of Congress Subject Headings (LCSH) [<a href="http://id.loc.gov/authorities/subjects.html">http://id.loc.gov/authorities/subjects.html</a>], Art and Architecture Thesaurus (AAT) [<a href="http://www.getty.edu/research/tools/vocabularies/index.html">http://www.getty.edu/research/tools/vocabularies/index.html</a>], and Thesaurus for Graphic Materials II (TGM II) [<a href="https://www.loc.gov/rr/print/tgm2/">https://www.loc.gov/rr/print/tgm2/</a>], while "sound recordings" appears in only the first two. Values given should be as specific as possible within the context of an institution's descriptive program. It is recommended that institutions adopt a consistent,</p>

	<p>well-documented approach to supplying genre terms.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/genre.html">https://www.loc.gov/standards/mods/userguide/genre.html</a>)</p> <p>»Dieses Element beschreibt die literarische Gattung eines Werkes. Es wird empfohlen, kontrolliertes Vokabular zu verwenden, das über die Attribute authority, authorityURI oder valueURI identifiziert wird. Mögliche Vokabulare sind:</p> <p>aad: Gattungsbegriffe der Arbeitsgemeinschaft Alter Drucke [<a href="http://aad.gbv.de/empfehlung/index.htm">http://aad.gbv.de/empfehlung/index.htm</a>]</p> <p>marcgt: MARC Genre Term List [<a href="http://www.loc.gov/standards/valuelist/marcgt.html">http://www.loc.gov/standards/valuelist/marcgt.html</a>] [...]</p> <p>Eines der folgenden Attribute kann verwendet werden:</p> <p>authority: Für den Namen des kontrollierten Vokabulars.</p> <p>authorityURI: Für den URI des kontrollierten Vokabulars.</p> <p>valueURI: Für den URI des Terms, der aus dem kontrollierten Vokabular verwendet wird.« (Rühle, Jahnke und Kühle, 2017, S. 14–15)</p>
Attribute	authority ; authorityURI ; valueURI
Konkordanzen	MARC 655 ; EAD <genreform> ; LIDO <objectWorkType> ; DCMES <type>
Filter	Objektgattung

<b>Element</b>	<subject>
Bezeichnung	Subject (englisch) ; Thema (deutsch)
URI	<a href="http://www.loc.gov/standards/mods/mods-outline-3-6.html#subject">http://www.loc.gov/standards/mods/mods-outline-3-6.html#subject</a>
Kardinalität	0..n
Definition	<p>»A term or phrase representing the primary topic(s) on which a work is focused.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/subject.html">https://www.loc.gov/standards/mods/userguide/subject.html</a>)</p>
Kommentar	<p>»&lt;subject&gt; is a container element that contains subelements relating to the subject aspect of the resource.</p> <p>If the subject is a controlled heading from an authoritative list, use the authority attribute to indicate which authority was used. If no authority attribute is present, the data is assumed to be an uncontrolled term.</p> <p>If it is not desired to parse the subject subelements, use &lt;subject&gt;&lt;topic&gt; and a string. [...]</p> <p>The DLF/Aquifer Implementation Guidelines for Shareable MODS Records require, when applicable, the use of at least one &lt;subject&gt; element in a record. Values for &lt;subject&gt; indicate what content is found within or represented by the work, and typically answer such questions as who, what, where, and when. [...]</p> <p>Whether or not the use of &lt;subject&gt; is applicable depends upon who might search for an item outside its local context and how they are likely to search for it. For instance, topical subject content may not apply to some items, such as abstract art. [...]</p> <p>DLF/Aquifer guidelines highly recommend the use of subject values from a controlled vocabulary or formal classification scheme and the designation of this list in the authority attribute. [...]</p>

	<p>Subelements within &lt;subject&gt; are used to differentiate subject content. While MODS does allow for placing multiple values in a single &lt;subject&gt;&lt;topic&gt; string, parsing subject terms into separate subelements is the preferred practice and highly recommended in the DLF/Aquifer guidelines. Place distinct, multiple subjects in separate &lt;subject&gt; fields. [...]</p> <p>Use of the authority attribute is recommended if applicable to record the name of the authoritative list for a controlled value. Specify authority at the &lt;subject&gt; level in most cases [...]. If providing subjects from different authorities, use a separate &lt;subject&gt; element for each.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/subject.html">https://www.loc.gov/standards/mods/userguide/subject.html</a>)</p> <p>»Für Schlagwörter gilt:</p> <p>Handelt es sich um kontrolliertes Vokabular, wird der Name des Vokabulars im Attribut authority benannt bzw. über authorityURI abgebildet. Wird das Schlagwort über einen valueURI identifiziert, kann auf die Angabe in authority verzichtet werden.</p> <p>Innerhalb von mods:subject können durch die Kombination von Unterelementen mehrere Schlagwörter kombiniert werden, falls sie jeweils demselben oder aber keinem kontrollierten Vokabular entstammen. [...]</p> <p>Der Name des kontrollierten Vokabulars bzw. der authorityURI wird im entsprechenden Attribut des Elements mods:subject genannt. Wird ein valueURI verwendet, so wird dieser im entsprechenden Attribut des jeweiligen Unterelements (z. B. mods:topic) angegeben. [...]</p> <p>Attribute:</p> <p>authority wenn das Schlagwort einer Normdatei entnommen wurde,  authorityURI wenn das Schlagwort einer Normdatei entnommen wurde, die durch einen URI identifizierbar ist.« (Rühle, Jahnke und Kühle, 2017, S. 24–25)</p>
Attribute	authority; authorityURI
Unterelemente	<topic>
Filter	siehe MODS <topic>

<b>Element</b>	<topic>
Bezeichnung	Topic (englisch) ; Sachschlagwort (deutsch)
URI	<a href="http://www.loc.gov/standards/mods/mods-outline-3-6.html#subject">http://www.loc.gov/standards/mods/mods-outline-3-6.html#subject</a>
Kardinalität	0..n
Definition	<p>»Used as the tag for any topical subjects that are not appropriate in the &lt;geographic&gt;, &lt;temporal&gt;, &lt;titleInfo&gt;, &lt;name&gt;, &lt;genre&gt;, &lt;hierarchicalGeographic&gt;, or &lt;occupation&gt; subelements.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/subject.html#topic">https://www.loc.gov/standards/mods/userguide/subject.html#topic</a>)</p>
Kommentar	<p>»If the type of subject string is unknown, &lt;topic&gt; is used (since &lt;subject&gt; is a container element). If a controlled subject term is used, indicate authority using the authority attribute at the &lt;subject&gt; level.« (Library of Congress - Network Development and MARC Standards Office, 8. Januar 2018, <a href="https://www.loc.gov/standards/mods/userguide/subject.html#topic">https://www.loc.gov/standards/mods/userguide/subject.html#topic</a>)</p> <p>»Use this subelement to indicate any primary topical subjects that are not appropriate in the &lt;geographic&gt;, &lt;temporal&gt;, &lt;titleInfo&gt;, or &lt;name&gt; subelements. While it is highly</p>

	<p>recommended that subject values be parsed into subelements, they may also be listed as a string under &lt;subject&gt;&lt;topic&gt;. This subelement has no attributes.</p> <p>If a controlled subject term is used, indicate authority using the authority attribute at the &lt;subject&gt; level. Locally developed terms should be listed separately, with local indicated as the source using the authority attribute at the &lt;subject&gt; level. If the term is uncontrolled (for example, if it is a keyword from legacy records), do not use the authority attribute.« (DLF Aquifer Metadata Working Group, 2009, S. 55–56)</p> <p>»Die Attribute authority und authorityURI werden in mods:subject angegeben, das Attribut valueURI wird jedoch in mods:topic verwendet.</p> <p>Attribute: valueURI wenn das Schlagwort durch einen URI identifizierbar ist (z. B. GND-URI)« (Rühle, Jahnke und Kühle, 2017, S. 25)</p>
Attribute	authority; authorityURI; valueURI
Konkordanzen	MARC 650 ; EAD <subject> ; LIDO <subjectConcept> ; DCMES <subject>
Filter	Themenschlagwort

## Encoded Archival Description, Version EAD3

<b>Element</b>	<genreform>
Bezeichnung	Genre/Physical Characteristic (englisch) ; Genre- und Formbegriffe (deutsch)
URI	<a href="http://www.loc.gov/ead/EAD3taglib/index.html#elem-genreform">http://www.loc.gov/ead/EAD3taglib/index.html#elem-genreform</a>
Kardinalität	0..n
Definition	<p>»An element for encoding a genre or form of material.« (Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists und Library of Congress - Network Development and MARC Standards Office, 2016, <a href="http://www.loc.gov/ead/EAD3taglib/index.html#elem-genreform">http://www.loc.gov/ead/EAD3taglib/index.html#elem-genreform</a>)</p>
Kommentar	<p>»Description and Usage: &lt;genreform&gt; identifies the types of material being described by naming the style or technique of their intellectual content (genre), order of information or object function (form), and physical characteristics. Examples include: account books, architectural drawings, portraits, short stories, sound recordings, and videotapes. [...]</p> <p>Use &lt;genreform&gt; within &lt;controlaccess&gt; for encoding genre terms as defined by controlled vocabularies or according to appropriate rules. You may also use &lt;genreform&gt; for encoding genre terms as they appear within text.« (Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists und Library of Congress - Network Development and MARC Standards Office, 2016, <a href="http://www.loc.gov/ead/EAD3taglib/index.html#elem-genreform">http://www.loc.gov/ead/EAD3taglib/index.html#elem-genreform</a>)</p> <p>»Beschreibung: Ein Begriff, der den Typ der erschlossenen Materialien kennzeichnet, indem die Ausführung oder das Herstellungsverfahren ihres Inhalts (Genre), der Informationsordnung oder der Objektfunktion (Form) und physische Eigenschaften benannt werden. Beispiele sind: Rechnungsbücher, Bauzeichnungen, Porträts, Kurzgeschichten, Tonaufnahmen und Videobänder.</p> <p>Nicht alle Genre und Formen der Materialien, die in einem Findbuch erwähnt sind, müssen getaggt werden. Eine Möglichkeit besteht darin, die Genre- und Formbegriffe zu taggen, für die eine andere Zugriffsmöglichkeit als eine einfache, undifferenzierte Volltextsuche gewünscht wird. Die Verwendung von kontrollierten Vokabularen wird empfohlen, um den Zugriff auf Informationen innerhalb und übergreifend in Verbundfindmitteln zu erleichtern. [...]</p>

	<p>Es sollte das Attribut SOURCE zur Angabe des Vokabulars verwendet werden, dem der Begriff entnommen wurde. Das Attribut NORMAL kann verwendet werden, um ansetzungskontrollierte Formen von Namen anzugeben, die in &lt;genreform&gt; in Form von Text, wie z. B. innerhalb eines Absatzes, erfasst wurden. Das Attribut AUTHFILENUMBER kann verwendet werden, um auf eine Normdatei zu verlinken, die mehr Informationen über den Namen oder Kreuzverweise zu alternativen Namensformen oder verwandten Namen gibt.« (Arbeitsgruppe Encoded Archival Description der Society of American Archivists und des Büros für Netzwerkentwicklung und MARC Standards der Library of Congress, 2006, S. 146)</p> <p>»Das normal-Attribut ist fakultativ. Es dient dazu den Archivalientyp in einer normierten Form abzubilden, um so dem Nutzer die Recherche zu erleichtern. Das normierte Vokabular umfasst folgende Kategorien: "Urkunden", "Siegel", "Amtsbücher [sic!]", Register und Grundbücher", "Akten", "Karten und Pläne", "Plakate und Flugblätter", "Drucksachen", "Bilder", "Handschriften", "Audio-Visuelle Medien", "Datenbanken", "Sonstiges"« (Landesarchiv Baden-Württemberg, 2013, S. 3)</p>
Attribute	authfilenumber ; normal ; source
Konkordanzen	MARC 655 ; MODS <genre> ; LIDO <objectWorkType> ; DCMES <type>
Filter	Objektgattung

Element	<subject>
Bezeichnung	Subject (englisch) ; Gegenstand (deutsch)
URI	<a href="http://www.loc.gov/ead/EAD3taglib/index.html#elem-subject">http://www.loc.gov/ead/EAD3taglib/index.html#elem-subject</a>
Kardinalität	0..n
Definition	<p>»An element for encoding topics represented in the materials being described.« (Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists und Library of Congress - Network Development and MARC Standards Office, 2016, <a href="http://www.loc.gov/ead/EAD3taglib/index.html#elem-subject">http://www.loc.gov/ead/EAD3taglib/index.html#elem-subject</a>)</p>
Kommentar	<p>»Description and Usage: Indicates a topic reflected in the described materials. [...] Use &lt;subject&gt; within &lt;controlaccess&gt; for encoding subjects as defined by controlled vocabularies or according to appropriate rules. You may also use &lt;subject&gt; for encoding subjects as they appear within text. [...]</p> <p>See also: Personal, corporate, family and geographic names used as subjects are tagged as &lt;persname&gt;, &lt;corpname&gt;, &lt;famname&gt;, and &lt;geogname&gt; respectively.« (Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists und Library of Congress - Network Development and MARC Standards Office, 2016, <a href="http://www.loc.gov/ead/EAD3taglib/index.html#elem-subject">http://www.loc.gov/ead/EAD3taglib/index.html#elem-subject</a>)</p> <p>»Beschreibung: Ein Begriff, der eine Sache, die in Verbindung mit den erschlossenen Materialien steht oder von ihnen behandelt wird, bezeichnet. Für Personennamen, Institutions- oder Ortsbezeichnungen, die als Gegenstand behandelt werden, wird &lt;persname&gt;, &lt;corpname&gt; und &lt;geogname&gt; verwendet. [...]</p> <p>Es müssen nicht alle Gegenstände zu einem Findbuch [sic!] kodiert werden. Eine Möglichkeit besteht darin, die Gegenstände zu taggen, für die eine andere Zugriffsmöglichkeit als eine einfache, undifferenzierte Volltextsuche gewünscht wird. Die Verwendung von Anwendungsformen wird empfohlen, um den Zugriff auf Gegenstandsbezeichnungen innerhalb und übergreifend in Verbundfindmitteln zu erleichtern. [...]</p> <p>Es sollte das Attribut SOURCE zur Angabe des Vokabulars verwendet werden, dem der Begriff entnommen wurde. Das Attribut NORMAL kann verwendet werden, um ansetzungskontrollierte Formen von Namen anzugeben, die in &lt;subject&gt; in Form von Text,</p>

	wie z. B. innerhalb eines Absatzes, erfasst wurden. [...] Das Attribut AUTHFILENUMBER kann verwendet werden, um auf eine Normdatei zu verlinken, die mehr Informationen über den Gegenstand oder Kreuzverweise zu alternativen Formen des Gegenstandsbegriffes gibt.« (Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists und Library of Congress - Network Development and MARC Standards Office, 2016, S. 229)
Attribute	authfilenumber ; normal ; source
Konkordanzen	MARC 650 ; MODS <topic> ; LIDO <subjectConcept> ; DCMES <subject>
Filter	Themenschlagwort

## Lightweight Information Describing Objects 1.0

Element	<classification>
Bezeichnung	Classification
URI	<a href="http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#classificationWrap">http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#classificationWrap</a>
Kardinalität	0..n
Definition	»Concepts used to categorize an object/work by grouping it together with others on the basis of similar characteristics.« (Coburn et al., 2010, S. 31)
Kommentar	<p>»How to record: The category belongs to a systematic scheme (classification) which groups objects of similar characteristics according to uniform aspects. This grouping/classification [sic!] may be done according to material, form, shape, function, region of origin, cultural context, or historical or stylistic period. In addition to this systematic grouping it may also be done according to organizational divisions within a museum (e.g., according to the collection structure of a museum). If the object/work [sic!] is assigned to multiple classifications, repeat this element. Preferably taken from a published controlled vocabulary.« (Coburn et al., 2010, S. 31–32)</p> <p>»Advice to implementers: An item can be classified in many ways (e.g. style, form, age, and phase), all of which can be represented in a LIDO record. It is also possible for multiple classifications in the same type (e.g. many colours).</p> <p>It is recommended that the term in a Classification comes from a controlled vocabulary. Include a URI if the term has been published online.</p> <p>The organization should be aware where Classification information is held in its collections management system in order to facilitate the creation of the LIDO record.« (McKenna und Stein, 2015, S. 39)</p> <p>»Gattung (empfohlen) [...] Tipps für die Erfassung: Dieses Datenfeld kann ergänzend zum Sachbegriff [Object/Work Type] verwendet werden. Manchmal gilt derselbe Begriff für beide Felder, z.B. Zeichnung, in der Regel umschreibt der Gattungsbegriff jedoch die Einordnung des Objekts in einen größeren Kontext. [...] Empfohlene Normdateien: lokale kontrollierte Werteliste, AAT, GND [...]</p> <p>Stilistische Einordnung [...] Empfohlene Normdateien: lokale kontrollierte Werteliste, AAT, GND« (Fichtl et al., 2015, S. 17-18)</p>
Attribute	type
Unterelemente	<conceptID> ; <term>

Konkordanzen	MARC 082 ; MARC 083 ; MARC 084 ; MODS <classification>
Filter	Sachkategorie

Element	<objectWorkType>
Bezeichnung	Object/Work Type [sic!]
URI	<a href="http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#objectWorkTypeWrap">http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#objectWorkTypeWrap</a>
Kardinalität	1..n
Definition	»The specific kind of object/work [sic!] being described.« (Coburn et al., 2010, S. 77)
Kommentar	<p>»How to record: Preferably taken from a published controlled vocabulary. For a collection, include repeating instances for identifying all of or the most important items in the collection.« (Coburn et al., 2010, S. 77–78)</p> <p>»Advice to implementers: It is mandatory to provide an object/work type in a LIDO record. Every LIDO record must have a term in this element. It is recommended that the term comes from a controlled vocabulary. Include a URI if the term has been published online. The organization must ensure that it has this information its collections management system in order to facilitate the creation of the LIDO record.« (McKenna und Stein, 2015, S. 38)</p> <p>»Sachbegriff/Objektbezeichnung (Pflichtelement) [...] Tipps für die Erfassung: Achtung: Dieses Datenfeld ist neben dem Titel des Werks das einzige inhaltliche LIDO Pflichtfeld und muss immer ausgefüllt werden. Das zu beschreibende Objekt wird hier mit einem oder mehreren Begriffen kurz charakterisiert. [...] Die Begriffe können abgeleitet sein von der Herstellungstechnik, dem ursprünglichen Funktionszusammenhang oder der Form (siehe Beispiele). Bitte in diesem Feld wann immer möglich die entsprechende ID des Begriffs aus dem Art and Architecture Thesaurus angeben (<a href="http://aat-deutsch.de">http://aat-deutsch.de</a> oder <a href="http://www.getty.edu/research/tools/vocabularies/aat">http://www.getty.edu/research/tools/vocabularies/aat</a>). Wenn mit IDs aus der GND gearbeitet worden ist, diese bitte vor dem Export mit dem AAT mappen, denn nur die Sammlung selbst kann entscheiden, welcher Term aus dem AAT genau dem aus der GND gewählten Deskriptor entspricht (Hier bitte stets darauf achten, dass die in den Normdateien genannten Oberbegriffe zu einander passen.). Es reicht nicht, nur die ID zu nennen, der Begriff muss explizit genannt werden. [...]</p> <p>Empfohlene Normdaten: lokale kontrollierte Werteliste, AAT, GND« (Fichtl et al., 2015, S: 13-14)</p>
Attribute	type
Unterelemente	<conceptID> ; <term>
Konkordanzen	MARC 655 ; MODS <genre> ; EAD <genreform> ; DCMES <type>
Filter	Objektgattung

Element	<subject>
Bezeichnung	Subject

URI	<a href="http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#subjectComplexType">http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#subjectComplexType</a>
Kardinalität	0..n
Definition	»Contains sub-elements for a structured subject description. These identify, describe, and/or interpret what is depicted in and by an object/work [sic!] or what it is about.« (Coburn et al., 2010, S. 110)
Attribute	type
Unterelemente	<subjectConcept>
Filter	siehe <subjectConcept>

<b>Element</b>	<subjectConcept>
Bezeichnung	Subject Concept
URI	<a href="http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#subjectComplexType">http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#subjectComplexType</a>
Kardinalität	0..n
Definition	»Provides references to concepts related to the subject of the described object/work [sic!].« (Coburn et al., 2010, S. 110)(Coburn et al., 2010, S. 110)
Kommentar	<p>»How to record: May include iconography, themes from literature, or generic terms describing the material world, or topics (e.g., concepts, themes, or issues). However, references to people, dates, events, places, objects are indicated in the [sic!] respective sub-elements Subject Actor Set, Subject Date Set, Subject Event Set, Subject Place Set, and Subject Object Set. Preferably taken from a published controlled vocabulary.« (Coburn et al., 2010, S. 110)</p> <p>»Ikongrafie / Dargestelltes Thema (empfohlen) [...] Empfohlene Normdateien: Iconclass, GND, AAT [...] Tipps für den LIDO-Export: Beim Export von Daten mit Iconclass-Notationen sollte nach Möglichkeit eine Ergänzung durch die den alphanumerischen Notationen zugeordneten Erläuterungen und Schlagwörter dieses hierarchischen Klassifikationssystems vorgenommen werden, um Bildthemen für die Recherche auch ohne Iconclass-Vorkenntnisse zugänglich zu machen. Das Element subjectConcept bietet für diesen Fall die Unterelemente conceptID (Iconclass-Notation) und term (Schlagwörter).« (Fichtl et al., 2015, S. 53-54)</p>
Unterelemente	<conceptID> ; <term>
Konkordanzen	MARC 650 ; MODS <topic> ; EAD <subject> ; DCMES <subject>
Filter	Themenschlagwort

<b>Element</b>	<conceptID>
Bezeichnung	Concept Identifier
URI	<a href="http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#identifierComplexType">http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#identifierComplexType</a>



Kardinalität	0..n
Definition	»A unique identifier for the concept.« (Coburn et al., 2010, S. 33)
Kommentar	<p>»How to record: Preferably taken from and linking to a published controlled vocabulary.« (Coburn et al., 2010, S. 33)</p> <p>»Sofern zu einem LIDO-Element, eine „ID“ (Concept, Place oder Actor [sic!]) vorhanden ist, kann diese für die Bildung von Links verwendet werden. Das Attribut „lido:source“ definiert die Herkunft der ID. Gebräuchliche Abkürzungen für verwendete Normdaten sind:</p> <p>&lt;lido:source=“AAT”&gt;   &lt;lido:source=“GND”&gt;   &lt;lido:source=“TGN”&gt;  </p> <p>&lt;lido:source=“Geonames”&gt; [sic!]</p> <p>Sofern eine ID nicht als URL angegeben ist, muss der lido:type=“local“ verwendet werden. Wenn „lido:source“ einer der folgenden Abkürzungen für Normdateien entspricht, wird aus der ID eine URL gebildet. Folgende Präfixe sind implementiert:</p> <p>AAT=http://vocab.getty.edu/aat/  TGN= http://vocab.getty.edu/aat/  VIAF=http://viaf.org/viaf/  GND=http://d-nb.info/gnd/  Iconclass=http://iconclass.org/  Geonames=http://www.geonames.org/  [...]</p> <p>Es wird empfohlen, die vollständige URL anzugeben und den lido:type=“URI“ zu verwenden.« (Fachstelle Museum der Deutschen Digitalen Bibliothek im Institut für Museumsforschung, 2014, S. 1)</p>
Attribute	type ; source

Element	<term>
Bezeichnung	Term/Label [sic!]
URI	<a href="http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#termComplexType">http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html#termComplexType</a>
Kardinalität	0..n
Definition	»A name for the referred concept, used for indexing.« (Coburn et al., 2010, S. 114)

## Dublin Core Metadata Element Set, Version 1.1

Element	<type>
Bezeichnung	Type (englisch) ; Typ (deutsch)
URI	<a href="http://purl.org/dc/elements/1.1/type">http://purl.org/dc/elements/1.1/type</a>
Kardinalität	0..n
Definition	<p>»The nature or genre of the resource.« (Dublin Core Metadata Initiative, 14. Juni 2012)</p> <p>»Art oder Gattung der Ressource.« (Kompetenzzentrum Interoperable Metadaten, 22. August 2007, S. 7)</p>

Kommentar	<p>»Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMI TYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.« (Dublin Core Metadata Initiative, 14. Juni 2012)</p> <p>»Generell wird empfohlen, ein kontrolliertes Vokabular zu verwenden wie zum Beispiel das DCMI Type Vocabulary. Das Dateiformat, der Träger und der Umfang der Ressource werden mit dem Element Format beschrieben.« (Kompetenzzentrum Interoperable Metadaten, 22. August 2007, S. 7)</p> <p>»Europeana note: The type of the original analog or born digital object as recorded by the content holder, this element typically includes values such as photograph, painting, sculpture etc. Although the portal needs normalised values to support type-related functions it is desirable to keep the original local values as well. Thus, all these original values should be mapped to this element. A separate europeana:type element has been added to contain the normalised value.« (Clayphan, 14. Juli 2013, S. 27–28)</p>
Konkordanzen	MARC 655 ; MODS <genre> ; EAD <genreform> ; LIDO <objectWorkType>
Filter	Objektgattung

<b>Element</b>	<subject>
Bezeichnung	Subject (englisch) ; Thema (deutsch)
URI	<a href="http://purl.org/dc/elements/1.1/subject">http://purl.org/dc/elements/1.1/subject</a>
Kardinalität	0..n
Definition	<p>»The topic of the resource.« (Dublin Core Metadata Initiative, 14. Juni 2012)</p> <p>»Das Thema der Ressource.« (Kompetenzzentrum Interoperable Metadaten, 22. August 2007, S. 7)</p>
Kommentar	<p>»Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.« (Dublin Core Metadata Initiative, 14. Juni 2012)</p> <p>»Üblicherweise wird das Thema durch Stichwörter, Schlagwörter oder Notationen beschrieben. Generell wird empfohlen, ein kontrolliertes Vokabular zu verwenden. Das räumliche oder zeitliche Thema der Ressource wird mit dem Element Geltungsbereich beschrieben.« (Kompetenzzentrum Interoperable Metadaten, 22. August 2007, S. 7)</p>
Konkordanzen	MARC 650 ; MODS <topic> ; EAD <subject> ; LIDO <subjectConcept> ; MARC 082 ; MARC 083 ; MARC 084 ; MODS <classification>
Filter	Themenschlagwort ; Sachkategorie

## Anhang B: Übersicht der in der DDB verwendeten kontrollierten Vokabulare

Nachfolgende Tabelle enthält eine Übersicht der von den Datengebern der DDB verwendeten kontrollierten Vokabulare, die im Rahmen der Stichprobenanalyse erhoben werden konnten.

In der Spalte »Vokabular« stehen Abkürzungen für folgende Bezeichnungen:

- AAD = Gattungs- und Sachbegriffe der Arbeitsgemeinschaft alter Drucke
- AAT = Art & Architecture Thesaurus
- DDC = Dewey Decimal Classification
- DDC-SG = DDC-Sachgruppen
- DIN 31631-4 = DIN 31631-4 Kriterienkatalog Dokumente; Codes für Einträge zu Datenkategorien
- DINI-Typen = Gemeinsames Vokabular für Publikations- und Dokumenttypen der DINI-AG Elektronisches Publizieren
- DNBGT = Liste der in der Deutschen Nationalbibliothek verwendeten Gattungsbegriffe
- EAD-Typ = Werteliste für den normierten Archivalientyp im EAD(DDB)-Profil
- GND-S = GND-Systematik
- LCSH = Library of Congress Subject Headings
- MARCGT = MARC Genre Term List
- OBG = Oberbegriffsdatei
- RVK = Regensburger Verbundklassifikation
- SSG = Sondersammelgebietsnummer

Die Notationen in der ersten Spalte stehen für folgende Gruppierungen:

- I = Klassifikationen und Systematiken
- II = Thesauri und Thesaurusartige
- III = Schlagwortlisten, Wertelisten und Nomenklaturen

Innerhalb dieser Gruppen sind die Vokabulare nach absteigender Anzahl der jeweiligen Indexterme geordnet. Die Spalte »LOD« gibt an, ob das Vokabular als Linked Open Data publiziert ist oder nicht. In der Spalte »Format« ist angegeben in welchem Metadatenformat bzw. in welchen Metadatenformaten das jeweilige Vokabular in den untersuchten Lieferdaten der DDB ausgewiesen ist.

Notation	Vokabular	Vokabulartyp	Inhaltliche Abdeckung	LOD	Indexterme (insgesamt: 47.171)				Format
					Insgesamt	in Prozent	mit URI	in Prozent	
I.1	DDC-SG	Systematik	Universell (Bibliothek)	Nein	4.286	9,1 %	0	0,0 %	MAR C
I.2	DDC	Klassifikation	Universell (Bibliothek)	Nein	1.075	2,3 %	0	0,0 %	MAR C, MOD S
I.3	Iconclass	Klassifikation	Ikonografie	Ja	648	1,4 %	504	1,1 %	LIDO, DCM ES
I.4	GND-S	Systematik	Universell (Bibliothek)	Ja	582	1,2 %	0	0,0 %	MAR C
I.5	SSG	Systematik	Fachgebiete (Bibliothek)	Nein	235	0,5 %	0	0,0 %	MAR C
I.6	RVK	Klassifikation	Universell (Bibliothek)	Nein	134	0,3 %	0	0,0 %	MAR C
II.1	GND	Thesaurusartige Struktur	Universell (Bibliothek)	Ja	3.314	7,0 %	2.724	5,8 %	MAR C, MOD S, LIDO
II.2	Partage	Thesaurus	Jugendstil-Objekte	Ja	286	0,6 %	286	0,6 %	LIDO
II.3	AAT	Thesaurus	Kunst und Architektur	Ja	250	0,5 %	230	0,5 %	LIDO
III.1	EAD-Typ	Werteliste	Archivalientypen	Nein	10.628	22,5 %	0	0,0 %	EAD
III.2	AAD	Schlagwortliste	Gattungsbegriffe, Fachgebiete	Geplant	388	0,8 %	0	0,0 %	MAR C, MOD S
III.3	LCSH	Nomenklatur	Universell (Bibliothek)	Ja	64	0,1 %	0	0,0 %	LIDO

III.4	MARCGT	Schlagwortliste	Dokumenttypen	Nein	54	0,1 %	0	0,0 %	MODS
III.5	DNBGT	Schlagwortliste	Gattungsbegriffe	Nein	21	0,0 %	0	0,0 %	MARC
III.6	Grobsystematik	Schlagwortliste	Objekttypen	Nein	19	0,0 %	0	0,0 %	LIDO
Notation	Vokabular	Vokabulartyp	Inhaltliche Abdeckung	LOD	Indexterme (insgesamt: 47.171)				Format
					Insgesamt	in Prozent	mit URI	in Prozent	
III.7	OBG	Schlagwortliste	Objekttypen	Ja	19	0,0 %	4	0,0 %	LIDO
III.8	DINI-Typen	Schlagwortliste	Dokumenttypen	Nein	6	0,0 %	0	0,0 %	DCMES
III.9	DIN 31631-4	Schlagwortliste	Dokumenttypen	Nein	4	0,0 %	0	0,0 %	MODS
					<b>22.013</b>	<b>46,7 %</b>	<b>3.748</b>	<b>7,9 %</b>	